Diffusion approximations and control variates for MCMC

Nicolas Brosse, Alain Durmus, Sean Meyn, Éric Moulines

Diffusion approximations and control variates for MCMC

Nicolas Brosse [1], Alain Durmus [2], Sean Meyn [3] and Éric Moulines [1]

November 25, 2018

Abstract

A new methodology is presented for the construction of control variates to reduce the variance of additive functionals of Markov Chain Monte Carlo (MCMC) samplers. Our control variates are defined as linear combinations of functions whose coefficients are obtained by minimizing a proxy for the asymptotic variance. The construction is theoretically justified by two new results. We first show that the asymptotic variances of some well-known MCMC algorithms, including the Random Walk Metropolis and the (Metropolis) Unadjusted/Adjusted Langevin Algorithm, are close to the asymptotic variance of the Langevin diffusion. Second, we provide an explicit representation of the optimal coefficients minimizing the asymptotic variance of the Langevin diffusion. Several examples of Bayesian inference problems demonstrate that the corresponding reduction in the variance is significant, and that in some cases it can be dramatic.

Keywords: Bayesian inference; Control variates; Langevin diffusion; Markov Chain Monte Carlo; Poisson equation; Variance reduction

1 Introduction

Let $U : \mathbb{R}^d \to \mathbb{R}$ be a measurable function on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\int_{\mathbb{R}^d} \mathrm{e}^{-U(x)} \mathrm{d}x < \infty$. This function is associated to a probability measure π on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ defined for all $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ by $\pi(\mathrm{A}) \stackrel{\mathrm{def}}{=} \int_{\mathrm{A}} \mathrm{e}^{-U(x)} \mathrm{d}x / \int_{\mathbb{R}^d} \mathrm{e}^{-U(x)} \mathrm{d}x$. We are interested in approximating $\pi(f) \stackrel{\mathrm{def}}{=} \int f(x)\pi(\mathrm{d}x)$, where f is a π-integrable function. The classical Monte Carlo solution to this problem is to simulate i.i.d. random variables $(X_k)_{k\in\mathbb{N}}$ with distribution π, and then to estimate $\pi(f)$ by the sample mean

$$\hat{\pi}_n(f) = n^{-1} \sum_{i=0}^{n-1} f(X_i) \ . \tag{1}$$

[1]Centre de Mathématiques Appliquées, UMR 7641, Ecole Polytechnique, France.
Emails: nicolas.brosse@polytechnique.edu, eric.moulines@polytechnique.edu

[2]Ecole Normale Supérieure CMLA 61, Av. du Président Wilson 94235 Cachan Cedex, France
Email: alain.durmus@cmla.ens-cachan.fr

[3]University of Florida, Department of Electrical and Computer Engineering, Gainesville, Florida.
Email: meyn@ece.ufl.edu

In most applications, sampling from π is not an option. Markov Chain Monte Carlo (MCMC) methods amount to sample a Markov chain $(X_k)_{k\in\mathbb{N}}$ from a Markov kernel R with (unique) invariant distribution π. Under weak additional conditions [MT09, Chapter 17], the estimator $\hat{\pi}_n(f)$ defined by (1) satisfies for any initial distribution a Central Limit Theorem (CLT)

$$n^{-1/2}\sum_{k=0}^{n-1}(f(X_k)-\pi(f)) \underset{n\to+\infty}{\overset{\text{weakly}}{\Longrightarrow}} \mathcal{N}(0,\sigma^2_{\infty,\mathrm{d}}(f)) \;, \quad \sigma^2_{\infty,\mathrm{d}}(f)=\pi\left((\hat{f}_{\mathrm{d}})^2-(R\hat{f}_{\mathrm{d}})^2\right) \;, \tag{2}$$

where $\mathcal{N}(m,\sigma^2)$ denotes a Gaussian distribution with mean m and variance σ^2, and $\hat{f}_{\mathrm{d}}$ is a solution of the Poisson equation

$$(R-\mathrm{Id})\hat{f}_{\mathrm{d}} = -\{f-\pi(f)\} \;. \tag{3}$$

Reducing the variance of Monte Carlo estimators is a very active research domain: see e.g. [RC04, Chapter 4], [Liu08, Section 2.3], and [RK17, Chapter 5] for an overview of the main methods. In this paper, we use control variates, *i.e.* π-integrable functions $h=(h_1,\ldots,h_p):\mathbb{R}^d\to\mathbb{R}^p$ satisfying $\pi(h_i)=0$ for $i\in\{1,\ldots,p\}$ and then choose $\theta\in\mathbb{R}^p$ such that $\sigma^2_{\infty,\mathrm{d}}(f+\theta^{\mathrm{T}}h)\leq\sigma^2_{\infty,\mathrm{d}}(f)$. [Hen97] and [Mey08, Section 11.5] proposed control variates of the form $(R-\mathrm{Id})\theta^{\mathrm{T}}\psi$ where $\psi=(\psi_1,\ldots,\psi_p)$ are known π-integrable functions. The parameter $\theta\in\mathbb{R}^p$ is obtained by minimizing the asymptotic variance

$$\min_{\theta\in\mathbb{R}^p}\sigma^2_{\infty,\mathrm{d}}(f+(R-\mathrm{Id})\theta^{\mathrm{T}}\psi)=\min_{\theta\in\mathbb{R}^p}\pi\left(\left\{\hat{f}_{\mathrm{d}}-\theta^{\mathrm{T}}\psi\right\}^2-\left\{R(\hat{f}_{\mathrm{d}}-\theta^{\mathrm{T}}\psi)\right\}^2\right) \;, \tag{4}$$

noting that $(-\theta^{\mathrm{T}}\psi)$ is a solution of the Poisson equation associated to $(R-\mathrm{Id})\theta^{\mathrm{T}}\psi$ and $\hat{f}_{\mathrm{d}}$ is defined in (3). The method suggested in [Mey08, Section 11.5] to minimize (4) requires estimates of the solution $\hat{f}_{\mathrm{d}}$ of the Poisson equation. Temporal Difference learning is a possible candidate, but this method is complex and suffers from high variance.

[DK12] noticed that if R is reversible w.r.t. π, it is possible to optimize the limiting variance (4) without computing explicitly the Poisson solution $\hat{f}_{\mathrm{d}}$. Reversibility will play an important role in this paper as well.

Each of the algorithms in the aforementioned literature requires computation of $R\psi_i$ for each $i\in\{1,\ldots,p\}$, which is in general a computational challenge. In [Hen97; Mey08] this is addressed by restricting to kernels for which $R(x,\cdot)$ has finite support for each x, and in [DK12] the authors restrict mainly to Gibbs samplers in their numerical examples.

In this paper an alternative class of control variates is used to avoid this computational barrier. This approach follows [AC99] (applications to quantum Monte Carlo calculations) and [MSI13; PMG14] (Bayesian statistics): assume that U is continuously differentiable, and for any twice continuously differentiable function φ, define $\mathscr{L}\varphi$ by

$$\mathscr{L}\varphi = -\langle\nabla U,\nabla\varphi\rangle+\Delta\varphi \;. \tag{5}$$

Under mild conditions on φ, it may be shown that $\pi(\mathscr{L}\varphi)=0$. [MSI13] suggested to use $\mathscr{L}(\theta^{\mathrm{T}}\psi)$ with $\psi=(\psi_1,\ldots,\psi_p)$ as control variates and choose θ by minimizing

$\theta \mapsto \pi(\{f-\pi(f)+\mathscr{L}\theta^{\mathrm{T}}\psi\}^2)$. This approach has triggered numerous work, among others [OGC16], [OG16] and [Oat+18] which introduce control functionals; a nonparametric extension of control variates. A drawback of this method stems from the fact that the optimization criterion $\pi(\{f-\pi(f)+\mathscr{L}\theta^{\mathrm{T}}\psi\}^2)$ is only theoretically justified if $(X_k)_{k\in\mathbb{N}}$ is i.i.d. and might significantly differ from the asymptotic variance $\sigma^2_{\infty,\mathrm{d}}(f+\mathscr{L}(\theta^{\mathrm{T}}\psi))$ defined in (1).

In this paper, we propose a new method to construct control variates. Analysis and motivation are based on the overdamped Langevin diffusion defined for $t \geq 0$ by

$$\mathrm{d}Y_t = -\nabla U(Y_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \;, \tag{6}$$

where $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion. If ∇U is Lipschitz, the Stochastic Differential Equation (SDE) (6) has a unique strong solution $(Y_t)_{t\geq 0}$ for every initial condition $x \in \mathbb{R}^d$. Denote by $(P_t)_{t\geq 0}$ the semigroup associated to the SDE (6) defined by $P_t f(x) = \mathbb{E}\left[f(Y_t)\right]$ where f is bounded measurable and $(Y_t)_{t\geq 0}$ is a solution of (6) started at x. Under mild additional conditions (see e.g. [RT96]), π is invariant for the semigroup $(P_t)_{t\geq 0}$, *i.e.* $\pi P_t = \pi$ for all $t \geq 0$. In addition, under smoothness and 'tail' conditions on f and ∇U, the following CLT holds for any initial condition (see [Bha82; CCG12])

$$t^{-1/2}\int_0^t \{f(Y_s)-\pi(f)\}\mathrm{d}s \overset{\text{weakly}}{\underset{t\to+\infty}{\Longrightarrow}} \mathcal{N}(0,\sigma^2_\infty(f)) \;, \quad \sigma^2_\infty(f) = 2\pi\left(\hat{f}\{f-\pi(f)\}\right) \;, \tag{7}$$

where $\hat{f} : \mathbb{R}^d \to \mathbb{R}$ is a solution of the (continuous-time) Poisson equation

$$\mathscr{L}\hat{f} = -\{f-\pi(f)\} \;. \tag{8}$$

The main contribution of this paper is the introduction of a new class of control variates based on the expression of the asymptotic variance $\sigma^2_\infty(f)$ given in (7). Since $\pi(\mathscr{L}(\theta^{\mathrm{T}}\psi)) = 0$ for any $\theta \in \mathbb{R}^d$, we consider the control variate $\mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi)$ where $\theta^*(f)$ is chosen by minimizing

$$\theta \mapsto \sigma^2_\infty(f+\mathscr{L}(\theta^{\mathrm{T}}\psi)) \;. \tag{9}$$

Although $\mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi)$ is a control variate for the Langevin diffusion associated with f, the choice of this optimization criterion is motivated by the fact that for some MCMC algorithms, the asymptotic variance $\sigma^2_{\infty,\mathrm{d}}(f)$ defined in (2) is (up to a scaling factor) a good approximation of the asymptotic variance of the Langevin diffusion $\sigma^2_\infty(f)$ defined in (7). Moreover, the minimization of (9) admits a unique solution $\theta^*(f)$, which is in general easy to estimate. It is worthwhile to note that it is not required to know the Poisson solution $\hat{f}$ to minimize (9).

The construction of control variates for MCMC and the related problem of approximating solutions of Poisson equations are very active fields of research. It is impossible to give credit for all the contributions undertaken in this area; see [DK12], [PMG14] and references therein for further background.

Amongst recent studies on this subject, [MV15] approximate directly the solution $\hat{f}_{\mathrm{d}}$ of the Poisson equation by subdividing the state space. Close to the methodology presented in the present paper, [MV17] uses the scaling limit of the RWM algorithm when the dimension d of the state space $\mathbb{R}^d$ goes to infinity to implement a control variates based on a solution of the Poisson equation for the Langevin diffusion. This approach uses a strong assumption on the stationary distribution which is assumed to be in product form. It is difficult to predict the performance of this methodology when this assumption is not met. Concerning the link between $\sigma^2_{\infty,\mathrm{d}}(f)$ and $\sigma^2_{\infty}(f)$, an analogous result associated to the error estimates for the Green-Kubo formula can be found in [LS16, Theorem 5.6].

The remainder paper is organized as follows. In Section 2, we present our methodology to compute the minimizer $\theta^*(f)$ of (9) and the construction of control variates for some MCMC algorithms. In Section 3, we state our main result which guarantees that the asymptotic variance $\sigma^2_{\infty,\mathrm{d}}(f)$ defined in (2) and associated with a given MCMC method is close (up to a scaling factor) to the asymptotic variance of the Langevin diffusion $\sigma^2_{\infty}(f)$ defined in (7). We provide a CLT and we show that under appropriate conditions on U, the Unadjusted Langevin Algorithm (ULA) fits the framework of our methodology. In Section 4, a Monte Carlo experiment illustrating the performance of our method is presented. In Section 5, we establish conditions under which the results of Sections 2 and 3 can be applied to the Random Walk Metropolis (RWM) and the Metropolis Adjusted Langevin Algorithm (MALA). The proofs are postponed to Section 6 and to the Appendix.

Notation

Let $\mathcal{B}(\mathbb{R}^d)$ denote the Borel σ-field of $\mathbb{R}^d$. Moreover, let $\mathrm{L}^1(\mu)$ be the set of μ-integrable functions for μ a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Further, $\mu(f) = \int_{\mathbb{R}^d} f(x)\mathrm{d}\mu(x)$ for an $f \in \mathrm{L}^1(\mu)$. Given a Markov kernel R on $\mathbb{R}^d$, for all $x \in \mathbb{R}^d$ and f integrable under $R(x,\cdot)$, denote by $Rf(x) = \int_{\mathbb{R}^d} f(y)R(x,\mathrm{d}y)$. Let $V : \mathbb{R}^d \to [1,\infty)$ be a measurable function. The V-total variation distance between two probability measures μ and ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is defined as $\|\mu - \nu\|_V = \sup_{|f|\le V} |\mu(f) - \nu(f)|$. If $V = 1$, then $\|\cdot\|_V$ is the total variation denoted by $\|\cdot\|_{\mathrm{TV}}$. For a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, define $\|f\|_V = \sup_{x\in\mathbb{R}^d} |f(x)|/V(x)$.

For $u, v \in \mathbb{R}^d$, define the scalar product $\langle u, v\rangle = \sum_{i=1}^d u_i v_i$ and the Euclidian norm $\|u\| = \langle u, u\rangle^{1/2}$. Denote by $\mathbb{S}(\mathbb{R}^d) = \{u \in \mathbb{R}^d : \|u\| = 1\}$. For $a, b \in \mathbb{R}$, denote by $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ and $a_+ = a \vee 0$. For $a \in \mathbb{R}_+$, $\lfloor a\rfloor$ and $\lceil a\rceil$ denote respectively the floor and ceil functions evaluated in a. We take the convention that for $n, p \in \mathbb{N}$, $n < p$ then $\sum_p^n = 0$, $\prod_p^n = 1$ and $\{p, \ldots, n\} = \emptyset$. Define for $t \in \mathbb{R}$, $\Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^t \mathrm{e}^{-r^2/2}\mathrm{d}r$ and $\bar\Phi(t) = 1 - \Phi(t)$.

For $k \in \mathbb{N}$, $m, m' \in \mathbb{N}^*$ and Ω, Ω' two open sets of $\mathbb{R}^m, \mathbb{R}^{m'}$ respectively, denote by $\mathrm{C}^k(\Omega, \Omega')$, the set of k-times continuously differentiable functions. For $f \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R})$, denote by ∇f the gradient of f and by Δf the Laplacian of f. For $k \in \mathbb{N}$ and $f \in \mathrm{C}^k(\mathbb{R}^d, \mathbb{R})$, denote by $\mathrm{D}^i f$ the i-th order differential of f for $i \in \{0, \ldots, k\}$. For $x \in \mathbb{R}^d$ and $i \in \{1, \ldots, k\}$, define $\|\mathrm{D}^0 f(x)\| = |f(x)|$, $\|\mathrm{D}^i f(x)\| = \sup_{u_1,\ldots,u_i\in\mathbb{S}(\mathbb{R}^d)} \mathrm{D}^i f(x)[u_1, \ldots, u_i]$.

For $k, p \in \mathbb{N}$ and $f \in \mathrm{C}^k(\mathbb{R}^d, \mathbb{R})$, define the norm

$$\|f\|_{k,p} = \sup_{x \in \mathbb{R}^d,\, i \in \{0,\ldots,k\}} \left\| \mathrm{D}^i f(x) \right\| / (1 + \|x\|^p) \,,$$

and $\mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}) = \left\{ f \in \mathrm{C}^k(\mathbb{R}^d, \mathbb{R}) : \inf_{p \in \mathbb{N}} \|f\|_{k,p} < +\infty \right\}$.

2 Langevin-based control variates for MCMC methods

Before introducing our new methodology based on the Langevin diffusion (6), we need to briefly recall some of its properties; this requires 'tail' and regularity assumptions on U. Let $k \geq 2$.

H1 (k)**.** *$U \in \mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and there exist $\upsilon > 0$ and $M_\upsilon \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq M_\upsilon$, $\langle \nabla U(x), x \rangle \geq \upsilon \|x\|$.*

Proposition 1. *Assume **H**1(k) for $k \geq 2$.*

(i) The semigroup $(P_t)_{t \geq 0}$ associated to (6) admits π as the unique invariant probability measure and for all $p \in \mathbb{N}$, $\int_{\mathbb{R}^d} \|x\|^p \, \pi(\mathrm{d}x) < +\infty$.

(ii) For any initial condition Y_0 and $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, the solution $(Y_t)_{t \geq 0}$ of the Langevin diffusion (6) satisfies the CLT (7).

(iii) For all $f \in \mathrm{C}^{k-1}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, there exists $\hat{f} \in \mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ such that $\mathscr{L}\hat{f} = \pi(f) - f$, where $\mathscr{L}$ is the generator of the Langevin diffusion defined in (5). For all $p \in \mathbb{N}$, there exist $C \geq 0$, $q \in \mathbb{N}$ such that for all $f \in \mathrm{C}^{k-1}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\|\hat{f}\|_{k,q} \leq C \|f\|_{k-1,p}$.

(iv) For all $f, g \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$,

$$\pi\left(f(-\mathscr{L})g\right) = \pi\left(g(-\mathscr{L})f\right) = \pi\left(\langle \nabla f, \nabla g \rangle\right) . \tag{10}$$

Proof. All these results are classical. A sketch of proof together with relevant references is postponed to Appendix C.1. $\square$

Proposition 1-(iii) ensures the existence and regularity of a solution of the Poisson equation (8) for any $f \in \mathrm{C}^{k-1}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $k \in \mathbb{N}$, $k \geq 2$. Proposition 1-(iv) is a classical "carré du champ" identity, see for example [BGL14, Section 1.6.2, formula 1.6.3]. It means in particular that the generator $\mathscr{L}$ is (formally) self-adjoint in $\mathrm{L}^2(\pi)$ which plays a key role in the construction of our control variates.

A straightforward consequence of (10) (setting $f = 1$) is that for any function $g \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\pi(\mathscr{L}g) = 0$. This suggests taking as a class of control variates for π the family of functions $\{\mathscr{L}(\theta^{\mathrm{T}}\psi) : \theta \in \mathbb{R}^p\}$, where $\psi = (\psi_1, \ldots, \psi_p) : \mathbb{R}^d \to \mathbb{R}^p$, $p \in \mathbb{N}^*$, is a fixed sieve of functions such that for all $i \in \{1, \ldots, p\}$, $\psi_i \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. Let $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$; by Proposition 1-(ii), for all $\theta \in \mathbb{R}^p$,

$$t^{-1/2} \int_0^t \{(f + \mathscr{L}(\theta^{\mathrm{T}}\psi))(Y_s) - \pi(f)\} \mathrm{d}s \overset{\text{weakly}}{\underset{t \to +\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_\infty(f + \mathscr{L}(\theta^{\mathrm{T}}\psi))) \,,$$

and an appropriate choice for the parameter $\theta \in \mathbb{R}^p$ is given by a minimizer of $\theta \mapsto \sigma^2_\infty(f + \mathscr{L}(\theta^{\mathrm{T}}\psi))$ defined in (7). We now show that this minimization problem has a unique solution which can be computed explicitly.

By Proposition 1-(iii), for any $f \in \mathrm{C}^1_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, the Poisson equation $\mathscr{L}\hat{f} = -\{f - \pi(f)\}$ has a solution $\hat{f} \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. Then, for all $\theta \in \mathbb{R}^p$, $\hat{f}_\theta = \hat{f} - \theta^{\mathrm{T}}\psi \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ is a solution of the Poisson equation $\mathscr{L}\hat{f}_\theta = -\{f_\theta - \pi(f_\theta)\}$, where $f_\theta = f + \mathscr{L}(\theta^{\mathrm{T}}\psi)$. Using the expression (7) of the asymptotic variance and $\pi(\mathscr{L}(\theta^{\mathrm{T}}\psi)) = 0$, we get for all $\theta \in \mathbb{R}^p$

$$\sigma^2_\infty(f_\theta) = 2\pi\left((\hat{f} - \theta^{\mathrm{T}}\psi)\left\{f - \pi(f) + \mathscr{L}(\theta^{\mathrm{T}}\psi)\right\}\right) . \tag{11}$$

Now by Proposition 1-(iv) and since $\mathscr{L}\hat{f} = \pi(f) - f$, we obtain

$$\pi(\hat{f}\mathscr{L}\psi) = \pi(\{\pi(f) - f\}\psi) .$$

Plugging this identity in (11) and using Proposition 1-(iv) imply for all $\theta \in \mathbb{R}^p$,

$$\sigma^2_\infty(f + \mathscr{L}(\theta^{\mathrm{T}}\psi)) = 2\theta^{\mathrm{T}}H\theta - 4\theta^{\mathrm{T}}\pi\left(\psi\left\{f - \pi(f)\right\}\right) + 2\pi\left(\hat{f}\left\{f - \pi(f)\right\}\right) ,$$

where $H \in \mathbb{R}^{p\times p}$ is given for any $i, j \in \{1, \ldots, p\}$ by

$$H_{ij} = \pi(\langle\nabla\psi_i, \nabla\psi_j\rangle) . \tag{12}$$

Therefore, $\theta \mapsto \sigma^2_\infty(f + \mathscr{L}(\theta^{\mathrm{T}}\psi))$ is a quadratic function and has a unique minimizer if and only if H is symmetric positive definite and this minimizer is given by

$$\theta^*(f) = H^{-1}\pi\left(\psi\{f - \pi(f)\}\right) . \tag{13}$$

Note that H is by definition a symmetric semi-positive definite matrix. It is easily seen that if $(1, \psi_1, \ldots, \psi_p)$ is linearly independent in $\mathrm{C}(\mathbb{R}^d, \mathbb{R})$, then H is full rank and the minimizer of $\sigma^2_\infty(f + \mathscr{L}(\theta^{\mathrm{T}}\psi))$ is given by (13).

To sum up, constructing control variates of the form $\mathscr{L}(\theta^{\mathrm{T}}\psi)$ for the Langevin diffusion is straightforward and the optimal parameter $\theta^*(f)$ minimizing the asymptotic variance has an explicit expression (13) that does not involve the (usually unknown) solution $\hat{f}$ of the Poisson equation $\mathscr{L}\hat{f} = \pi(f) - f$.

Implications for MCMC The continuous-time setting has mainly theoretical interest. The main contribution of this paper is to show that the optimal control variate for the diffusion remains nearly optimal for many classes of discrete-time MCMC algorithms.

One example is the Markov kernel associated with the Unadjusted Langevin Algorithm (ULA). A diffusion approximation is to be expected since the ULA algorithm is the Euler discretization scheme associated to the Langevin SDE (6): $X_{k+1} = X_k - \gamma\nabla U(X_k) + \sqrt{2\gamma}Z_{k+1}$, where $\gamma > 0$ is the step size and $(Z_k)_{k\in\mathbb{N}}$ is an i.i.d. sequence of standard Gaussian d-dimensional random vectors. The idea of using the Markov chain $(X_k)_{k\in\mathbb{N}}$ to sample approximately from π has been first introduced in the physics literature by [Par81] and popularized in the computational statistics community by [Gre83] and [GM94]. Other examples include the Metropolis Adjusted Langevin Algorithm (MALA) algorithm, and the Random Walk Metropolis algorithm (RWM).

Each of these MCMC algorithms define a family of Markov kernels $\{R_\gamma, \gamma \in (0,\bar{\gamma}]\}$, indexed by the step-size parameter $\gamma \in (0,\bar{\gamma}]$, for $\bar{\gamma} > 0$. For any initial distribution ξ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $\gamma \in (0,\bar{\gamma}]$, denote by $\mathbb{P}_{\xi,\gamma}$ and $\mathbb{E}_{\xi,\gamma}$ the probability and the expectation respectively on the canonical space of a Markov chain with initial distribution ξ and of transition kernel R_γ. By convention, we set $\mathbb{E}_{x,\gamma} = \mathbb{E}_{\delta_x,\gamma}$ for all $x \in \mathbb{R}^d$. We denote by $(X_k)_{k\geq 0}$ the canonical process. Under $\mathbb{P}_{\xi,\gamma}$, $(X_k)_{k\geq 0}$ is a Markov chain with initial distribution ξ and Markov kernel R_γ. The following assumptions are imposed here. General criteria to justify (I)–(III) are postponed to Section 3.

(I) For each $\gamma \in (0,\bar{\gamma}]$, R_γ is a positive Harris Markov kernel with invariant distribution π_γ satisfying $\pi_\gamma(|f|) < \infty$ for any $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$.

(II) For any initial condition X_0, each $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\gamma \in (0,\bar{\gamma}]$,

$$\sqrt{n}(\hat{\pi}_n(f) - \pi_\gamma(f)) \underset{n\to+\infty}{\overset{\text{weakly}}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_{\infty,\gamma}(f)) \tag{14}$$

where $\hat{\pi}_n(f)$ is defined by (1), and $\sigma^2_{\infty,\gamma}(f) \geq 0$ is the asymptotic variance associated with f defined by (2) relatively to R_γ.

(III) For any functions f, g sufficiently smooth and satisfying growth conditions,

$$\sigma^2_{\infty,\gamma}(f + \gamma g) = \gamma^{-1}\sigma^2_\infty(f) + o(\gamma^{-1}) \tag{15}$$

$$\pi_\gamma(f) = \pi(f) + O(\gamma) \,, \tag{16}$$

where $\sigma^2_\infty(f)$ is defined in (7) and for $\gamma \downarrow 0^+$.

The standard conditions (I)–(II) are in particular satisfied if R_γ is V-uniformly geometrically ergodic, see e.g. [MT09]. Let $V : \mathbb{R}^d \to [1,+\infty)$ be a measurable function. We say that R_γ, $\gamma \in (0,\bar{\gamma}]$ is V-uniformly geometrically ergodic if it admits an invariant probability measure π_γ such that $\pi_\gamma(V) < +\infty$ and there exist $C \geq 0$ and $\rho \in [0,1)$ such that for any probability measure ξ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $n \in \mathbb{N}$,

$$\|\xi R^n_\gamma - \pi_\gamma\|_V \leq C\xi(V)\rho^n \,.$$

The approximation result (III) requires more sophisticated arguments given in Section 3.

Based on (I)–(III) and (13), the estimator of $\pi(f)$ we suggest is given for $n \in \mathbb{N}^*$ by

$$\pi^{\mathrm{CV}}_n(f) = \frac{1}{n}\sum_{k=0}^{n-1}\left\{f(X_k) + \mathscr{L}\left(\hat{\theta}^*_n(f)^{\mathrm{T}}\psi(X_k)\right)\right\} \,, \tag{17}$$

and $\hat{\theta}^*_n(f)$ is an estimator of $\theta^*(f)$ defined in (13) given by

$$\hat{\theta}^*_n(f) = H^+_n\left[\frac{1}{n}\sum_{k=0}^{n-1}\psi(X_k)\left\{f(X_k) - \hat{\pi}_n(f)\right\}\right] \,, \tag{18}$$

where H_n^+ is the Moore-Penrose pseudoinverse of $H_n \in \mathbb{R}^{p\times p}$ defined for all $i,j \in \{1,\ldots,p\}$ by

$$(H_n)_{ij} = \frac{1}{n}\sum_{k=0}^{n-1} \langle \nabla\psi_i(X_k), \nabla\psi_j(X_k)\rangle \ . \tag{19}$$

We sketch informally the arguments required to justify (17). Since, under (I), for any $\gamma \in (0,\bar{\gamma}]$, the Markov kernel R_γ is positive Harris, by the strong law of large numbers, $\hat{\pi}_n(\psi\{f-\hat{\pi}_n(f)\})$ and H_n converge $\mathbb{P}_{\xi,\gamma}$-almost surely for any initial probability measure ξ to $\pi_\gamma(\{f-\pi_\gamma(f)\}\psi)$ and H_γ where

$$(H_\gamma)_{ij} = \pi_\gamma\left(\langle\nabla\psi_i, \nabla\psi_j\rangle\right) \ , \quad i,j \in \{1,\ldots,p\} \ . \tag{20}$$

If $(1,\psi_1,\ldots,\psi_p)$ is linearly independent in $\mathrm{C}(\mathbb{R}^d,\mathbb{R})$ and π_γ admits a positive density w.r.t. the Lebesgue measure, H_γ is a symmetric positive definite matrix. Hence, the sequence $(\hat{\theta}_n^*(f))_{n\geq 0}$ converges $\mathbb{P}_{\xi,\gamma}$-almost surely to

$$\theta_\gamma^*(f) = H_\gamma^{-1}\pi_\gamma\left\{(f-\pi_\gamma(f))\psi\right\} \ . \tag{21}$$

Under (II), using standard arguments, see Proposition 15, the following central limit theorem holds

$$\sqrt{n\gamma}\left\{\pi_n^{\mathrm{CV}}(f) - \pi_\gamma(f + \mathscr{L}(\theta_\gamma^*(f)^{\mathrm{T}}\psi))\right\} \overset{\mathbb{P}_{\xi,\gamma}-\text{weakly}}{\underset{n\to+\infty}{\Longrightarrow}} \mathcal{N}(0, \gamma\sigma^2_{\infty,\gamma}(f + \mathscr{L}(\theta_\gamma^*(f)^{\mathrm{T}}\psi))) \ . \tag{22}$$

Moreover, under (III), since $\theta_\gamma^*(f) = \theta^*(f) + O(\gamma)$, we get that

$$\begin{aligned}\gamma\sigma^2_{\infty,\gamma}(f + \mathscr{L}(\theta_\gamma^*(f)^{\mathrm{T}}\psi)) &= \gamma\sigma^2_{\infty,\gamma}\left(f + \mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi) + O(\gamma)\sum_{i=1}^p \mathscr{L}\psi_i\right) \\ &= \sigma^2_\infty(f + \mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi)) + o(1) \ ,\end{aligned} \tag{23}$$

for $\gamma \downarrow 0^+$. Therefore for any $\gamma \in (0,\bar{\gamma}]$, we get

$$\sqrt{n\gamma}\left\{\pi_n^{\mathrm{CV}}(f) - \pi_\gamma(f + \mathscr{L}(\theta_\gamma^*(f)^{\mathrm{T}}\psi))\right\} \overset{\mathbb{P}_{\xi,\gamma}-\text{weakly}}{\underset{n\to+\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_\infty(f + \mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi)) + o(1)) \ , \tag{24}$$

showing that the optimal control variate for the Langevin diffusion $\mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi)$ is asymptotically optimal as $\gamma \downarrow 0^+$ for the considered MCMC algorithm. Note that (24) also displays the existence of a bias term $\pi_\gamma(f + \mathscr{L}(\theta_\gamma^*(f)^{\mathrm{T}}\psi)) - \pi(f)$ which vanishes when $\pi_\gamma = \pi$. As shown in Section 3, we may get rid of the bias term by letting the step size γ depend on the number of samples n.

3 Asymptotic expansion for the asymptotic variance of MCMC algorithms

In this Section, we justify (III). Let $\bar{\gamma} > 0$, $V : \mathbb{R}^d \to [1,+\infty)$ and $k \in \mathbb{N}$. Consider the following assumptions:

A 1 $(V, \bar{\gamma})$. *There exist $\lambda \in [0,1)$, $b < +\infty$ and $c > 0$ such that*

$$\sup_{x \in \mathbb{R}^d} \{\exp(c\|x\|)/V(x)\} < +\infty \quad \text{and} \quad R_\gamma V \leq \lambda^\gamma V + \gamma b \,, \quad \text{for all } \gamma \in (0, \bar{\gamma}]. \tag{25}$$

Moreover, there exists $\varepsilon \in (0,1]$ such that for all $\gamma \in (0, \bar{\gamma}]$ and $x, x' \in \{V \leq M\}$,

$$\|R_\gamma^{\lceil 1/\gamma \rceil}(x, \cdot) - R_\gamma^{\lceil 1/\gamma \rceil}(x', \cdot)\|_{\mathrm{TV}} \leq 1 - \varepsilon \,, \tag{26}$$

where

$$M > \left(\frac{4b\lambda^{-\bar{\gamma}}}{\log(1/\lambda)} - 1 \right) \vee 1 \,. \tag{27}$$

A2 $(\bar{\gamma}, k)$. *There exist $\alpha \geq 3/2$ and a family of operators $(\mathscr{A}_\gamma)_{\gamma \in (0, \bar{\gamma}]}$ with $\mathscr{A}_\gamma : \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R}) \to \mathrm{C}^{i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ for $i \in \{0, \ldots, k\}$, such that for all $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\gamma \in (0, \bar{\gamma}]$,*

$$R_\gamma \varphi = \varphi + \gamma \mathscr{L} \varphi + \gamma^\alpha \mathscr{A}_\gamma \varphi \,.$$

For all $p \in \mathbb{N}$, there exist $C \geq 0$ and $q \in \mathbb{N}$ such that for all $i \in \{0, \ldots, k\}$, $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\gamma \in (0, \bar{\gamma}]$,

$$\|\mathscr{A}_\gamma \varphi\|_{i,q} \leq C \|\varphi\|_{4+i,p} \,.$$

For any $\varphi \in \mathrm{C}^{4}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $x \in \mathbb{R}^d$, $\lim_{\gamma \downarrow 0^+} \mathscr{A}_\gamma \varphi(x)$ exists (this limit is denoted $\mathscr{A}_0 \varphi(x)$).

We show below and in Section 5 that these conditions are satisfied for the Metropolis Adjusted / Unadjusted Langevin Algorithm (MALA and ULA) algorithms (in which case γ is the stepsize in the Euler discretization of the Langevin diffusion) and also by the Random Walk Metropolis algorithm (RWM) (in which case γ is the variance of the increment distribution). In the following result, we establish the V-uniform geometric ergodicity of R_γ for $\gamma \in (0, \bar{\gamma}]$.

Lemma 2. *Let $V : \mathbb{R}^d \to [1, +\infty)$ and $\bar{\gamma} > 0$. Assume **A** 1$(V, \bar{\gamma})$. For all $\gamma \in (0, \bar{\gamma}]$, R_γ has a unique invariant measure π_γ. There exist $C > 0$ and $\rho \in (0,1)$ such that for all $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$,*

$$\|\delta_x R_\gamma^n - \pi_\gamma\|_V \leq C \rho^{n\gamma} V(x) \,. \tag{28}$$

Proof. The proof is postponed to Section 6.1. □

Note that under **A** 1$(V, \bar{\gamma})$, iterating the drift condition (25), using Lemma 2 and $1 - \lambda^\gamma \geq \gamma \ln(1/\lambda) \lambda^\gamma$, we obtain for all $\gamma \in (0, \bar{\gamma}]$, $n \in \mathbb{N}^*$, $x \in \mathbb{R}^d$,

$$R_\gamma^n V(x) \leq \lambda^{n\gamma} V(x) + \frac{b\lambda^{-\bar{\gamma}}}{\ln(1/\lambda)} \quad \text{and} \quad \pi_\gamma(V) \leq \frac{b\lambda^{-\bar{\gamma}}}{\ln(1/\lambda)} \,. \tag{29}$$

We next give an upper bound on the bias between π_γ and π, *i.e.* $|\pi_\gamma(\varphi) - \pi(\varphi)|$ for φ smooth enough and $\pi_\gamma \neq \pi$.

Proposition 3. *Assume* $\boldsymbol{H}$1*(4). Let* $V : \mathbb{R}^d \to [1, +\infty)$, $\bar{\gamma} > 0$ *and assume* $\boldsymbol{A}$1*(*$V, \bar{\gamma}$*) and* $\boldsymbol{A}$2*(*$\bar{\gamma}, 0$*). For all* $p \in \mathbb{N}$*, there exists* $C \geq 0$ *such that for all* $\varphi \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ *and* $\gamma \in (0, \bar{\gamma}]$*,*

$$|\pi_\gamma(\varphi) - \pi(\varphi)| \leq C \, \|\varphi\|_{3,p} \, \gamma^{\alpha-1} \, . \tag{30}$$

Proof. The proof is postponed to Section 6.2. □

We now state the main theorem of this Section.

Theorem 4. *Let* $V : \mathbb{R}^d \to [1, +\infty)$, $\bar{\gamma} > 0$*. Assume* $\boldsymbol{H}$1*(7),* $\boldsymbol{A}$1*(*$V, \bar{\gamma}$*) and* $\boldsymbol{A}$2*(*$\bar{\gamma}, 3$*). Then, for all* $p \in \mathbb{N}$*, there exists* $C \geq 0$ *such that for all* $f \in \mathrm{C}^6_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$*, and* $n \in \mathbb{N}^*$

$$\left| \frac{\gamma}{n} \mathbb{E}_{x,\gamma} \left[\left(\sum_{k=0}^{n-1} \{ f(X_k) - \pi_\gamma(f) \} \right)^2 \right] - \sigma^2_\infty(f) \right| \leq C \, \|f\|^2_{6,p} \left\{ \gamma^{(\alpha-1)\wedge 1} + \frac{V(x)}{n\gamma} \right\} \, , \tag{31}$$

where $\sigma^2_\infty(f)$ *is defined in* (7)*.*

Proof. The proof is postponed to Section 6.3. □

Bias and confidence intervals In the CLT given in (24), the bias $\pi_\gamma(f + \mathscr{L}(\theta^*_\gamma(f)^{\mathrm{T}}\psi)) - \pi(f)$ is different from 0, except if $\pi_\gamma = \pi$. To obtain asymptotically valid confidence intervals for $\pi(f)$, we let the step size γ depend on the total number of iterations n.

Let $(\gamma_n)_{n \in \mathbb{N}^*}$ be a positive sequence and $\pi^{\mathrm{CV}}_{n,\gamma_n}(f)$ be defined in (17) where $(X_k)_{k \in \mathbb{N}}$ is associated to the kernel R_{γ_n}. We show that, for an appropriate sequence $(\gamma_n)_{n \in \mathbb{N}^*}$, $\pi^{\mathrm{CV}}_{n,\gamma_n}(f)$ targets $\pi(f)$ and a CLT holds with an asymptotic variance equal to $\sigma^2_\infty(f + \mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi))$. The optimal control variates for the Langevin diffusion $\mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi)$ is then also optimal for the MCMC algorithm of kernel R_{γ_n} in the limit $n \to +\infty$.

Theorem 5. *Let* $V : \mathbb{R}^d \to [1, +\infty)$, $\bar{\gamma} > 0$*. Assume* $\boldsymbol{H}$1*(10),* $\boldsymbol{A}$1*(*$V, \bar{\gamma}$*), and* $\boldsymbol{A}$2*(*$\bar{\gamma}, 6$*). Let* $f \in \mathrm{C}^9_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\psi = (\psi_1, \ldots, \psi_p) : \mathbb{R}^d \to \mathbb{R}^p$, $p \in \mathbb{N}^*$*, be a fixed sieve of functions such that* $(1, \psi_1, \ldots, \psi_p)$ *is linearly independent in* $\mathrm{C}(\mathbb{R}^d, \mathbb{R})$ *and for all* $i \in \{1, \ldots, p\}$, $\psi_i \in \mathrm{C}^{11}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$*. Let* $(\gamma_n)_{n \in \mathbb{N}^*}$ *be a positive sequence satisfying* $\lim_{n \to +\infty} (n\gamma_n)^{-1} + \gamma_n = 0$, $\hat{f}$ *be a solution of the Poisson equation* $\mathscr{L}\hat{f} = \pi(f) - f$, $\theta^*(f)$ *be defined in* (13) *and* ξ *be a probability measure such that* $\xi(V) < +\infty$*. Then,*

(i) if $\pi(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi)) \lim_{n \to +\infty} n^{1/2}\gamma_n^{\alpha-1/2} = 0$,

$$n^{1/2}\gamma_n^{1/2} \left\{ \pi^{\mathrm{CV}}_{n,\gamma_n}(f) - \pi(f) \right\} \overset{\mathbb{P}_{\xi,\gamma_n}-\text{weakly}}{\underset{n \to +\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_\infty(f + \mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi))) \, ,$$

(ii) if $\lim_{n \to +\infty} n^{1/2}\gamma_n^{\alpha-1/2} = \gamma_\infty \in [0, +\infty)$,

$$n^{1/2}\gamma_n^{1/2} \left\{ \pi^{\mathrm{CV}}_{n,\gamma_n}(f) - \pi(f) \right\} \overset{\mathbb{P}_{\xi,\gamma_n}-\text{weakly}}{\underset{n \to +\infty}{\Longrightarrow}} \mathcal{N}(\gamma_\infty \pi(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi)), \sigma^2_\infty(f + \mathscr{L}(\theta^*(f)^{\mathrm{T}}\psi))) \, ,$$

(iii) *if* $\pi(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi)) \liminf_{n\to+\infty} n^{1/2}\gamma_n^{\alpha-1/2} = +\infty$,

$$\gamma_n^{1-\alpha}\left\{\pi_{n,\gamma_n}^{\mathrm{CV}}(f) - \pi(f)\right\} \overset{\mathbb{P}_{\xi,\gamma_n}-weakly}{\underset{n\to+\infty}{\Longrightarrow}} \pi\left(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi)\right) \quad ,$$

where $\pi_{n,\gamma_n}^{\mathrm{CV}}(f)$ *and* $\sigma_\infty^2(f)$ *are defined in* (17) *and* (7)*, respectively.*

Proof. The proof is postponed to [Bro+18, Section B]. □

Note that if the invariant distribution of R_γ is π for all $\gamma \in (0, \bar{\gamma}]$ (e.g. the case of MALA or RWM), we have under **A**2$(\bar{\gamma}, 0)$ and by the dominated convergence theorem, $\pi(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi)) = 0$ and (i) always holds.

The ULA algorithm The Markov kernel R_γ^{ULA} associated to the ULA algorithm is given for $\gamma > 0$, $x \in \mathbb{R}^d$ and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$R_\gamma^{\mathrm{ULA}}(x, \mathrm{A}) = (4\pi\gamma)^{-d/2} \int_{\mathrm{A}} \exp\left(-(4\gamma)^{-1} \|y - x + \gamma\nabla U(x)\|^2\right) \mathrm{d}y \,. \tag{32}$$

Based on the results of [DM17] and [DM16], the following lemmas enable to check **A** 1$(V, \bar{\gamma})$ and **A**2$(\bar{\gamma}, k)$, $k \in \mathbb{N}$, for the ULA algorithm. Analysis of the MALA and RWM algorithms is postponed to Section 5. Consider the following assumptions on U.

H2. *$U : \mathbb{R}^d \to \mathbb{R}$ is gradient Lipschitz, i.e. there exists $L \geq 0$ such that for all $x, y \in \mathbb{R}^d$, $\|\nabla U(x) - \nabla U(y)\| \leq L\|x - y\|$.*

H3. *There exist $\nu > 0$, $\alpha \in (1, 2]$, a minimizer $x^\star \in \arg\min U$ and $M_\nu \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\nu$, $\langle \nabla U(x), x - x^\star \rangle \geq \nu \|x - x^\star\|^\alpha$.*

H4. *U is convex and admits a minimizer $x^\star \in \arg\min U$.*

For simplicity we have assumed that ∇U is Lipschitz but following [Bro+17], this assumption can be relaxed. Note that under **H**4, by [Bra+14, Lemma 2.2.1], there exist $\eta > 0$ and $M_\eta \geq 0$ such that for all $x \in \mathbb{R}^d$, $\|x - x^\star\| \geq M_\eta$, $\langle \nabla U(x), x - x^\star \rangle \geq U(x) - U(x^\star) \geq \eta \|x - x^\star\|$ where $x^\star \in \arg\min_{\mathbb{R}^d} U$. Let $k \in \mathbb{N}$, $k \geq 2$. Note that if $U \in \mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and U satisfies **H**2 and **H**3 or **H**4, then **H**1(k) holds.

Lemma 6. *(i) Assume **H**2. R_γ^{ULA} satisfies the Doeblin condition* (26)*.*

*(ii) Assume **H**2 and **H**3 or **H**4. Then **A** 1(V, L^{-1}) is satisfied where V is defined for $x \in \mathbb{R}^d$ by,*

$$V(x) = \begin{cases} \exp(U(x)/2) & \text{under } \boldsymbol{H}3 \\ \exp\left((\eta/4)(1 + \|x - x^\star\|^2)^{1/2}\right) & \text{under } \boldsymbol{H}4. \end{cases}$$

Proof. The proof is postponed to Appendix C.2. □

To establish $\mathbf{A}\,2(\bar\gamma, 6)$, let $i \in \{0, \ldots, 6\}$, $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\bar\gamma > 0$, $\gamma \in [0, \bar\gamma]$ and $x \in \mathbb{R}^d$. Using $X_1 = X_0 - \gamma \nabla U(X_0) + \sqrt{2\gamma} Z_1$ where Z_1 is an i.i.d. standard d-dimensional Gaussian vector, we get

$$
\begin{aligned}
&\varphi(X_1) = \varphi(x) - \gamma \langle \nabla U(x), \nabla \varphi(x) \rangle + \sqrt{2\gamma} \langle \nabla \varphi(x), Z \rangle + \gamma \, \mathrm{D}^2 \varphi(x)[Z^{\otimes 2}] \\
&- \sqrt{2} \gamma^{3/2} \, \mathrm{D}^2 \varphi(x)[\nabla U(x), Z] + (\gamma^2/2) \, \mathrm{D}^2 \varphi(x)[\nabla U(x)^{\otimes 2}] - \gamma^2 \, \mathrm{D}^3 \varphi(x)[\nabla U(x), Z^{\otimes 2}] \\
&- (1/6) \gamma^3 \, \mathrm{D}^3 \varphi(x)[\nabla U(x)^{\otimes 3}] + 2^{-1/2} \gamma^{5/2} \, \mathrm{D}^3 \varphi(x)[\nabla U(x)^{\otimes 2}, Z] \\
&+ \frac{\sqrt{2}}{3} \gamma^{3/2} \, \mathrm{D}^3 \varphi(x)[Z^{\otimes 3}] + \frac{1}{6} \int_0^1 (1-t)^3 \, \mathrm{D}^4 \varphi(x + t(X_1 - x))[(X_1 - x)^{\otimes 4}] \mathrm{d}t \,. \qquad (33)
\end{aligned}
$$

Taking the expectation in (33), we obtain $R_\gamma \varphi(x) = \varphi(x) + \gamma \mathscr{L} \varphi(x) + \gamma^2 \mathscr{A}^{\mathrm{ULA}}_\gamma \varphi(x)$ where,

$$
\begin{aligned}
&\mathscr{A}^{\mathrm{ULA}}_\gamma \varphi(x) = \frac{1}{2} \mathrm{D}^2 \varphi(x)[\nabla U(x)^{\otimes 2}] - \frac{1}{6} \gamma \, \mathrm{D}^3 \varphi(x)[\nabla U(x)^{\otimes 3}] - \mathbb{E}\left[\mathrm{D}^3 \varphi(x)[\nabla U(x), Z^{\otimes 2}]\right] \\
&+ \frac{1}{6} \int_0^1 (1-t)^3 \mathbb{E}\left[\mathrm{D}^4 \varphi(x - t\gamma \nabla U(x) + t\sqrt{2\gamma} Z)[(-\sqrt{\gamma} \nabla U(x) + \sqrt{2} Z)^{\otimes 4}]\right] \mathrm{d}t \,. \qquad (34)
\end{aligned}
$$

Taking the limit $\gamma \downarrow 0^+$ in (34), we get for any $\varphi \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $x \in \mathbb{R}^d$, $\lim_{\gamma \downarrow 0^+} \mathscr{A}^{\mathrm{ULA}}_\gamma \varphi(x) = \mathscr{A}^{\mathrm{ULA}}_0 \varphi(x)$ where

$$
\mathscr{A}^{\mathrm{ULA}}_0 \varphi(x) = \frac{1}{2} \mathrm{D}^2 \varphi(x)[\nabla U(x)^{\otimes 2}] - \mathbb{E}\left[\mathrm{D}^3 \varphi(x)[\nabla U(x), Z^{\otimes 2}]\right] + \frac{1}{6} \mathbb{E}\left[\mathrm{D}^4 \varphi(x)[Z^{\otimes 4}]\right] \,. \qquad (35)
$$

Summarizing this discussion, it is easy to show that

Lemma 7. *Assume that $U \in \mathrm{C}^7_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$.*

(i) For all $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $i \in \{0, \ldots, 6\}$, $\mathscr{A}^{\mathrm{ULA}}_\gamma \varphi \in \mathrm{C}^i_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ for $\gamma > 0$ and for any $\bar\gamma > 0$ and $p \in \mathbb{N}$, there exist $C \geq 0$, $q \in \mathbb{N}$ such that for all $i \in \{0, \ldots, 6\}$, $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\gamma \in (0, \bar\gamma]$, $\|\mathscr{A}^{\mathrm{ULA}}_\gamma \varphi\|_{i,q} \leq C \|\varphi\|_{4+i,p}$.

(ii) For any $\varphi \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $x \in \mathbb{R}^d$, $\lim_{\gamma \downarrow 0^+} \mathscr{A}^{\mathrm{ULA}}_\gamma \varphi(x) = \mathscr{A}^{\mathrm{ULA}}_0 \varphi(x)$.

If $U \in \mathrm{C}^7_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, under **H**2 and **H**3 or **H**4, by Lemma 6 and Lemma 7, $\mathbf{A}\,1(V, L^{-1})$ and $\mathbf{A}\,2(L^{-1}, 6)$ with $\alpha = 2$ are satisfied; Theorem 4 and Theorem 5 then hold.

4 Numerical experiments

We illustrate the proposed control variates method on Bayesian logistic and probit regressions, see [Gel+14, Chapter 16], [MR07, Chapter 4]. The examples and the data sets are taken from [PMG14]. The code used to run the experiments is available at https://github.com/nbrosse/controlvariates. Let $\mathsf{Y} = (\mathsf{Y}_1, \ldots, \mathsf{Y}_n) \in \{0,1\}^N$ be a vector of binary response variables, $x \in \mathbb{R}^d$ be the regression coefficients, and $\mathsf{X} \in \mathbb{R}^{N \times d}$

be a design matrix. The log-likelihood for the logistic and probit regressions are given respectively by

$$
\begin{aligned}
\ell_{\log}(\mathsf{Y}|x, \mathsf{X}) &= \sum_{i=1}^{N} \left\{ \mathsf{Y}_i \mathsf{X}_i^{\mathrm{T}} x - \ln\left(1 + \mathrm{e}^{\mathsf{X}_i^{\mathrm{T}} x}\right) \right\} , \\
\ell_{\mathrm{pro}}(\mathsf{Y}|x, \mathsf{X}) &= \sum_{i=1}^{N} \left\{ \mathsf{Y}_i \ln(\Phi(\mathsf{X}_i^{\mathrm{T}} x)) + (1 - \mathsf{Y}_i) \ln(\Phi(-\mathsf{X}_i^{\mathrm{T}} x)) \right\} ,
\end{aligned}
$$

where $\mathsf{X}_i^{\mathrm{T}}$ is the i^{th} row of X for $i \in \{1, \ldots, N\}$. For both models, a Gaussian prior of mean 0 and variance $\varsigma^2 \,\mathrm{Id}$ is assumed for x where $\varsigma^2 = 100$. The posterior probability distributions $\pi_{\log}$ and π_{pro} for the logistic and probit regressions are proportional for all $x \in \mathbb{R}^d$ to

$$
\begin{aligned}
\pi_{\log}(x|\mathsf{Y}, \mathsf{X}) &\propto \exp\left(-U_{\log}(x)\right) \quad \text{with} \quad U_{\log}(x) = -\ell_{\log}(\mathsf{Y}|x, \mathsf{X}) + (2\varsigma^2)^{-1} \left\|x\right\|^2 , \\
\pi_{\mathrm{pro}}(x|\mathsf{Y}, \mathsf{X}) &\propto \exp\left(-U_{\mathrm{pro}}(x)\right) \quad \text{with} \quad U_{\mathrm{pro}}(x) = -\ell_{\mathrm{pro}}(\mathsf{Y}|x, \mathsf{X}) + (2\varsigma^2)^{-1} \left\|x\right\|^2 .
\end{aligned}
$$

In the following lemma, we check the assumptions on $U_{\log}$ and U_{pro} in order to apply Theorem 4 and Theorem 5 for the ULA, MALA and RWM algorithms. Note that **H**5 and **H** 6 are two additional conditions on U given in Section 5, introduced to check **A**1$(\exp(U/2), \bar{\gamma})$ and **A**2$(\bar{\gamma}, 6)$, for the RWM algorithm.

Lemma 8. *$U_{\log}$ and U_{pro} satisfy **H**1(k) for any $k \in \mathbb{N}^*$, **H**2, **H**4, **H**5 and **H**6.*

Proof. The proof is postponed to Appendix C.3. □

Following [PMG14, Section 2.1], we compare two bases for the construction of a control variate, based on first and second degree polynomials. Define $\psi^{1\mathrm{st}} = (\psi_1^{1\mathrm{st}}, \ldots, \psi_d^{1\mathrm{st}})$ given for $i \in \{1, \ldots, d\}$ and $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ by $\psi_i^{1\mathrm{st}}(x) = x_i$ and $\psi^{2\mathrm{nd}} = (\psi_1^{2\mathrm{nd}}, \ldots, \psi_{d(d+3)/2}^{2\mathrm{nd}})$ given for $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ by

$$
\begin{aligned}
&\psi_k^{2\mathrm{nd}}(x) = x_k \quad \text{for } k \in \{1, \ldots, d\} , \quad \psi_{k+d}^{2\mathrm{nd}}(x) = x_k^2 \quad \text{for } k \in \{1, \ldots, d\} , \\
&\psi_k^{2\mathrm{nd}}(x) = x_i x_j \quad \text{for } k = 2d + (j-1)(d - j/2) + (i - j) \text{ and all } 1 \leq j < i \leq d .
\end{aligned}
$$

$\psi^{1\mathrm{st}}$ and $\psi^{2\mathrm{nd}}$ are in $\mathrm{C}^{\infty}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and are linearly independent in $\mathrm{C}(\mathbb{R}^d, \mathbb{R})$. The estimators associated to $\psi^{1\mathrm{st}}$ and $\psi^{2\mathrm{nd}}$ are referred to as CV-1 and CV-2, respectively.

For the ULA, MALA and RWM algorithms, we make a run of $n = 10^6$ samples with a burn-in period of 10^5 samples, started at the mode of the posterior. The step size is set equal to 10^{-2} for ULA and to 5×10^{-2} for MALA and RWM, the acceptance ratio in the stationary regime being close to 0.23 for RWM and 0.57 for MALA, see [RGG97; RR98]. We consider $2d$ scalar test functions $\{f_k\}_{k=1}^{2d}$ defined for all $x \in \mathbb{R}^d$ and $k \in \{1, \ldots, d\}$ by $f_k(x) = x_k$ and $f_{k+d}(x) = x_k^2$. For $k \in \{1, \ldots, 2d\}$, we compute the empirical average $\hat{\pi}_n(f_k)$ and the control variate estimator $\pi_{n,\gamma_n}^{\mathrm{CV}}(f_k)$ defined in (1) and (17) respectively. For comparison purposes, the zero-variance estimators of [PMG14] using the same bases of functions ψ^{1st}, ψ^{2nd} are also computed and are referred to as ZV-1 for ψ^{1st} and ZV-2 for ψ^{2nd}. We run 100 independent Markov chains for ULA, MALA, RWM algorithms. The boxplots for the logistic example are displayed in Figure 2 for x_1 and x_1^2. Note the impressive decrease in the variance using the control variates for each algorithm ULA, MALA and RWM. It is worthwhile to note that for ULA, the bias $|\pi(f) - \pi_\gamma(f)|$ is reduced dramatically using the CV-2 estimator. It can be explained by the fact that for n large enough, $\theta_n^*(f)^{\mathrm{T}}\psi^{\mathrm{2nd}}$ is an efficient approximation of the solution $\hat{f}$ of the Poisson equation $\mathscr{L}\hat{f} = -(f - \pi(f))$. We then get

$$\lim_{n\to+\infty} \pi_{n,\gamma_n}^{\mathrm{CV}}(f) \approx \pi_\gamma(f) + \pi_\gamma\left(\mathscr{L}(\theta_\gamma^*(f)^{\mathrm{T}}\psi^{\mathrm{2nd}})\right) \approx \pi_\gamma(f) - \pi_\gamma\left(f - \pi(f)\right) = \pi(f)$$

where $\pi_{n,\gamma_n}^{\mathrm{CV}}(f)$ is defined in (17).

To have a more quantitative estimate of the variance reduction, we compute for each algorithm and test function $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, the spectral estimator $\hat{\sigma}_n^2(f)$ of the asymptotic variance with a Tukey-Hanning window, see [FJ10], given by

$$\hat{\sigma}_n^2(f) = \sum_{k=-(\lfloor n^{1/2}\rfloor-1)}^{\lfloor n^{1/2}\rfloor-1} \left\{\frac{1}{2} + \frac{1}{2}\cos\left(\frac{\pi\,|k|}{\lfloor n^{1/2}\rfloor}\right)\right\}\omega_n^f(|k|)\;, \tag{36}$$

$$\omega_n^f(k) = \frac{1}{n}\sum_{s=0}^{n-1-k}\left\{f(X_s) - \hat{\pi}_n(f)\right\}\left\{f(X_{s+k}) - \hat{\pi}_n(f)\right\}\;.$$

We compute the average of these estimators $\hat{\sigma}_n^2(f)$ over the 100 independent runs of the Markov chains and the values for the logistic regression are given in Table 1. The Variance Reduction Factor (VRF) is defined as the ratio of the asymptotic variances obtained by the ordinary empirical average and the control variate (or zero-variance) estimator. We again observe the considerable decrease of the asymptotic variances using control variates. In this example, our approach produces slightly larger VRFs compared to the zero-variance estimators. We obtain similar results for the probit regression; see Appendix D.

5 The RWM and MALA algorithms

In this Section, we establish the assumptions of Theorem 4 and Theorem 5 for the RWM and MALA algorithms. For $\gamma > 0$, the Markov kernel R_γ^{RWM} of the RWM algorithm with a Gaussian proposal of mean 0 and variance 2γ is given for $x \in \mathbb{R}^d$ and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$\begin{aligned} R_\gamma^{\mathrm{RWM}}(x, \mathrm{A}) &= \int_{\mathrm{A}} \exp\left(-(4\gamma)^{-1}\|y - x\|^2\right)\min(1, \mathrm{e}^{-\tau^{\mathrm{RWM}}(x,y)})\frac{\mathrm{d}y}{(4\pi\gamma)^{d/2}} \\ &\quad + \delta_x(\mathrm{A})\left\{1 - \int_{\mathbb{R}^d}\exp\left(-(4\gamma)^{-1}\|y - x\|^2\right)\min(1, \mathrm{e}^{-\tau^{\mathrm{RWM}}(x,y)})\frac{\mathrm{d}y}{(4\pi\gamma)^{d/2}}\right\}\;, \end{aligned} \tag{37}$$

$$\tau^{\mathrm{RWM}}(x, y) = U(y) - U(x)\;. \tag{38}$$

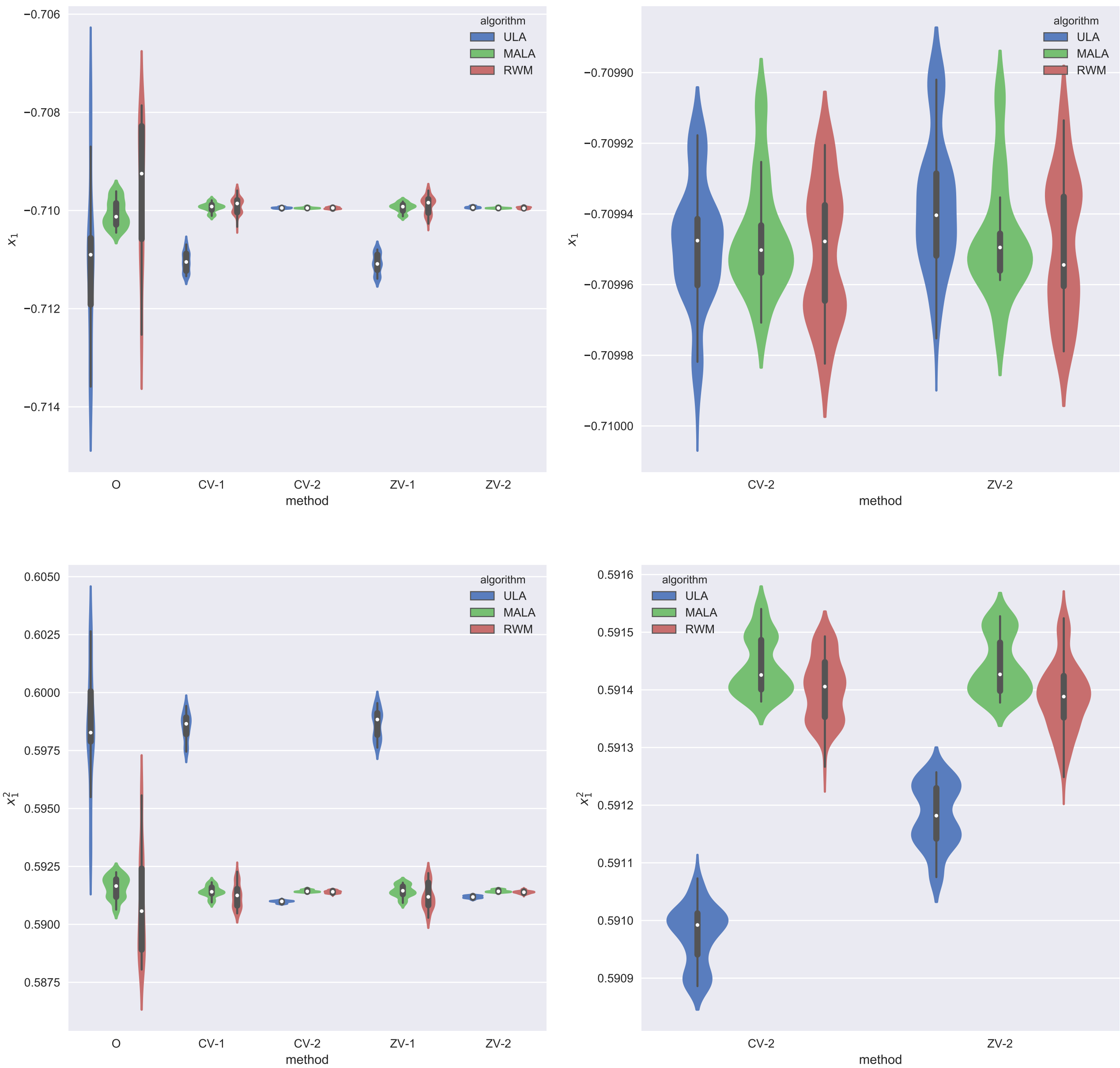

Figure 1: Boxplots of x_1, x_1^2 using the ULA, MALA and RWM algorithms for the logistic regression. The compared estimators are the ordinary empirical average (O), our estimator with a control variate (17) using first (CV-1) or second (CV-2) order polynomials for ψ, and the zero-variance estimators of [PMG14] using a first (ZV-1) or second (ZV-2) order polynomial bases. The plots in the second column are close-ups for CV-2 and ZV-2.

Table 1: Estimates of the asymptotic variances for ULA, MALA and RWM and each parameter x_i, x_i^2 for $i \in \{1, \dots, d\}$, and of the variance reduction factor (VRF) on the example of the logistic regression.

		MCMC	CV-1-MCMC		CV-2-MCMC		ZV-1-MCMC		ZV-2-MCMC	
		Variance	VRF	Variance	VRF	Variance	VRF	Variance	VRF	Variance
x_1	ULA	2	33	0.061	3.2e+03	0.00062	33	0.061	3e+03	0.00066
	MALA	0.41	33	0.012	2.6e+03	0.00016	30	0.014	2.5e+03	0.00017
	RWM	1.3	33	0.039	2.6e+03	0.00049	32	0.04	2.7e+03	0.00048
x_2	ULA	10	57	0.18	8.1e+03	0.0013	53	0.19	7.4e+03	0.0014
	MALA	2.5	59	0.042	7.7e+03	0.00032	54	0.046	7.3e+03	0.00034
	RWM	5.6	52	0.11	5.6e+03	0.001	50	0.11	5.6e+03	0.001
x_2	ULA	10	56	0.18	7.3e+03	0.0014	52	0.19	6.7e+03	0.0015
	MALA	2.4	58	0.041	6.8e+03	0.00035	52	0.045	6.5e+03	0.00037
	RWM	5.6	45	0.13	5.1e+03	0.0011	42	0.13	5.1e+03	0.0011
x_4	ULA	13	26	0.5	3.9e+03	0.0033	22	0.59	3.4e+03	0.0038
	MALA	3.1	25	0.12	3.6e+03	0.00087	21	0.14	3.3e+03	0.00095
	RWM	7.5	19	0.4	2.5e+03	0.003	18	0.43	2.4e+03	0.0031
x_1^2	ULA	4.6	10	0.46	5.5e+02	0.0084	9.3	0.49	4.8e+02	0.0095
	MALA	0.98	9.6	0.1	4.6e+02	0.0021	8.6	0.11	4.2e+02	0.0023
	RWM	3	8.3	0.36	4.3e+02	0.0069	8	0.37	4.3e+02	0.0069
x_2^2	ULA	29	11	2.6	5.2e+02	0.055	10	2.8	4.7e+02	0.062
	MALA	7	11	0.64	5.2e+02	0.013	10	0.68	4.8e+02	0.014
	RWM	16	9.1	1.8	4.4e+02	0.037	8.8	1.8	4.3e+02	0.037
x_3^2	ULA	46	11	4.1	6.7e+02	0.069	10	4.5	5.9e+02	0.079
	MALA	11	11	0.97	6e+02	0.018	10	1	5.6e+02	0.019
	RWM	26	9	2.9	4.3e+02	0.061	8.6	3.1	4.2e+02	0.062
x_4^2	ULA	5.1e+02	14	37	8.2e+02	0.62	12	43	6.9e+02	0.73
	MALA	1.2e+02	14	9	7.9e+02	0.15	12	10	7.1e+02	0.17
	RWM	2.9e+02	11	27	5.8e+02	0.51	10	29	5.6e+02	0.53

For all $x \in \mathbb{R}^d$ and $\gamma > 0$, define the acceptance region

$$\mathbf{A}_{x,\gamma}^{\mathrm{RWM}} = \left\{ z \in \mathbb{R}^d : \tau^{\mathrm{RWM}}(x, x + \sqrt{2\gamma} z) \leq 0 \right\} \tag{39}$$

and denote by $\partial \mathbf{A}_{x,\gamma}^{\mathrm{RWM}}$ the boundaries of the connected components of $\mathbf{A}_{x,\gamma}^{\mathrm{RWM}}$.

H5. *For all $x \in \mathbb{R}^d$ and $\gamma > 0$, $\partial \mathbf{A}_{x,\gamma}^{\mathrm{RWM}}$ is a Lebesgue null set.*

Set for all $\gamma > 0$,

$$\mathscr{A}_\gamma^{\mathrm{RWM}} = (R_\gamma^{\mathrm{RWM}} - \mathrm{Id} - \gamma \mathscr{L}) / \gamma^{3/2} \ . \tag{40}$$

If $U \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, define for any $\varphi \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $x, z \in \mathbb{R}^d$,

$$\begin{aligned}
\mathscr{A}_0^{\mathrm{RWM}} \varphi(x) &= -\sqrt{2} \mathbb{E} \left[\langle \nabla U(x), Z \rangle_+ \mathrm{D}^2 \varphi(x)[Z^{\otimes 2}] + \zeta_0(x, Z) \langle \nabla \varphi(x), Z \rangle \right] \ , \\
\zeta_0(x, z) &= \left\{ \mathrm{D}^2 U(x)[z^{\otimes 2}] + \langle \nabla U(x), z \rangle^2 \right\} \mathbb{1}_{\langle \nabla U(x), z \rangle > 0} \\
&\qquad + \left(\mathrm{D}^2 U(x)[z^{\otimes 2}] \right)_+ \mathbb{1}_{\langle \nabla U(x), z \rangle = 0} \ ,
\end{aligned} \tag{41}$$

where Z is a standard d-dimensional Gaussian vector.

Lemma 9. *(i) Assume that $U \in \mathrm{C}^7_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and **H**5. For all $i \in \{0, \ldots, 6\}$ and $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\mathscr{A}_\gamma^{\mathrm{RWM}} \varphi \in \mathrm{C}^i_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ for $\gamma > 0$ and for any $\bar{\gamma} > 0$ and $p \in \mathbb{N}$, there exist $C \geq 0$, $q \in \mathbb{N}$ such that for all $\varphi \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $\gamma \in (0, \bar{\gamma}]$, $\|\mathscr{A}_\gamma^{\mathrm{RWM}} \varphi\|_{0,q} \leq C \|\varphi\|_{4,p}$.*

(ii) Assume that $U \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. For any $\varphi \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $x \in \mathbb{R}^d$, $\lim_{\gamma \downarrow 0^+} \mathscr{A}_\gamma^{\mathrm{RWM}} \varphi(x) = \mathscr{A}_0^{\mathrm{RWM}} \varphi(x)$.

Proof. This result follows from [BDM18]. □

We now proceed to check the drift condition (25). In that purpose, consider the following additional assumption on U.

H6. *There exist $\chi, M > 0$ such that for all $x \in \mathbb{R}^d$, $\|x\| \geq M$,*

$$\|\nabla U(x)\| \geq \chi^{-1} \ , \quad \left\| \mathrm{D}^3 U(x) \right\| \leq \chi \left\| \mathrm{D}^2 U(x) \right\| \ , \quad \left\| \mathrm{D}^2 U(x) \right\| \leq \chi \|\nabla U(x)\|$$

and $\lim_{\|x\| \to +\infty} \left\| \mathrm{D}^2 U(x) \right\| / \|\nabla U(x)\|^2 = 0$.

Lemma 10. *Assume that $U \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and **H**6. There exists $\bar{\gamma} > 0$ such that for all $\gamma \in (0, \bar{\gamma}]$, R_γ^{RWM} satisfies the drift condition (25) with $V = \exp(U/2)$.*

Proof. This result follows from [BDM18]. □

We now consider the MALA algorithm. The Markov kernel R_γ^{MALA} of the MALA algorithm, see [RT96], is given for $\gamma > 0$, $x \in \mathbb{R}^d$, and $\mathrm{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$\begin{aligned} R_\gamma^{\mathrm{MALA}}(x, \mathrm{A}) &= \int_{\mathrm{A}} R_\gamma^{\mathrm{ULA}}(x, \mathrm{d}y) \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,y)}) \\ &\qquad + \delta_x(\mathrm{A}) \left\{ 1 - \int_{\mathbb{R}^d} R_\gamma^{\mathrm{ULA}}(x, \mathrm{d}y) \min(1, \mathrm{e}^{-\tau_\gamma^{\mathrm{MALA}}(x,y)}) \right\} , \qquad (42) \\ \tau_\gamma^{\mathrm{MALA}}(x, y) &= U(y) - U(x) + \frac{\|x - y + \gamma \nabla U(y)\|^2 - \|y - x + \gamma \nabla U(x)\|^2}{4\gamma} . \qquad (43) \end{aligned}$$

For all $x \in \mathbb{R}^d$ and $\gamma > 0$, define the acceptance region

$$\mathbf{A}_{x,\gamma}^{\mathrm{MALA}} = \left\{ z \in \mathbb{R}^d : \tau_\gamma^{\mathrm{MALA}}(x, x - \gamma \nabla U(x) + \sqrt{2\gamma} z) \le 0 \right\}$$

and denote by $\partial \mathbf{A}_{x,\gamma}^{\mathrm{MALA}}$ the boundaries of the connected components of $\mathbf{A}_{x,\gamma}^{\mathrm{MALA}}$.

H 7. *For all $x \in \mathbb{R}^d$ and $\gamma > 0$, $\partial \mathbf{A}_{x,\gamma}^{\mathrm{MALA}}$ is a Lebesgue null set.*

Under this assumption, the following Lemma shows that $\mathbf{A}2(\bar{\gamma}, k)$ is satisfied for the MALA algorithm for any $\bar{\gamma} > 0$ and $k \in \mathbb{N}$. Set for all $\gamma > 0$,

$$\mathscr{A}_\gamma^{\mathrm{MALA}} = (R_\gamma^{\mathrm{MALA}} - \mathrm{Id} - \gamma \mathscr{L}) / \gamma^2 , \qquad (44)$$

and define for all $\varphi \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $x, z \in \mathbb{R}^d$,

$$\begin{aligned} \mathscr{A}_0^{\mathrm{MALA}} \varphi(x) &= \mathscr{A}_0^{\mathrm{ULA}} \varphi(x) - \sqrt{2} \mathbb{E} \left[\max(0, \xi_0(x, Z)) \langle \nabla \varphi(x), Z \rangle \right] , \\ \xi_0(x, z) &= -(\sqrt{2}/6) \, \mathrm{D}^3 U(x)[z^{\otimes 3}] + 2^{-1/2} \left\langle \nabla U(x), \mathrm{D}^2 U(x)[z] \right\rangle , \qquad (45) \end{aligned}$$

where Z is a standard d-dimensional Gaussian vector and $\mathscr{A}_0^{\mathrm{ULA}} \varphi$ is given in (35).

Lemma 11. *(i) Assume that $U \in \mathrm{C}^{10}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and **H** 7. For all $i \in \{0, \ldots, 6\}$ and $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\mathscr{A}_\gamma^{\mathrm{MALA}} \varphi \in \mathrm{C}^i_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ for $\gamma > 0$ and for any $\bar{\gamma} > 0$ and $p \in \mathbb{N}$, there exist $C \ge 0$, $q \in \mathbb{N}$ such that for all $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $i \in \{0, \ldots, 6\}$ and $\gamma \in (0, \bar{\gamma}]$, $\|\mathscr{A}_\gamma^{\mathrm{MALA}} \varphi\|_{i,q} \le C \|\varphi\|_{4+i,p}$.*

(ii) Assume that $U \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. For any $\varphi \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $x \in \mathbb{R}^d$, $\lim_{\gamma \downarrow 0^+} \mathscr{A}_\gamma^{\mathrm{MALA}} \varphi(x) = \mathscr{A}_0^{\mathrm{MALA}} \varphi(x)$.

Proof. The proof is postponed to Appendix C.4. □

6 Proofs

6.1 Proof of Lemma 2

By (25), the drift condition $\mathrm{D}_g(V, \lambda^\gamma, \gamma b)$ is satisfied. By (26) and (27), the set $\{V \le M\}$ is an $(\lceil 1/\gamma \rceil, 1 - \varepsilon)$-Doeblin set and by the choice of M, see (27), $\lambda^\gamma + (2\gamma b)/(1 + M) < 1$. A direct application of [Dou+18, Theorem 18.4.3] concludes the proof.

6.2 Proof of Proposition 3

Under $\mathbf{H}1(4)$, by Proposition 1-(iii) with $k=4$, there exist $q_1 \in \mathbb{N}$ and $C \geq 0$ such that for all $\varphi \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\hat{\varphi} \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ where $\mathscr{L}\hat{\varphi} = \pi(\varphi) - \varphi$ and $\|\hat{\varphi}\|_{4,q_1} \leq C\|\varphi\|_{3,p}$. Under $\mathbf{A}2(\bar{\gamma}, 0)$, we have for all $\gamma \in (0, \bar{\gamma}]$,

$$R_\gamma \hat{\varphi} = \hat{\varphi} + \gamma \mathscr{L}\hat{\varphi} + \gamma^\alpha \mathscr{A}_\gamma \hat{\varphi} \; . \tag{46}$$

By Proposition 1-(iii), $\mathbf{A}2(\bar{\gamma}, 0)$ and (29), there exist $q_2 \in \mathbb{N}$ and $C \geq 0$ such that for all $\gamma \in (0, \bar{\gamma}]$,

$$|\pi_\gamma(\mathscr{A}_\gamma \hat{\varphi})| \leq \pi_\gamma(|\mathscr{A}_\gamma \hat{\varphi}|) \leq C \, \|\mathscr{A}_\gamma \hat{\varphi}\|_{0,q_2} \leq C \, \|\hat{\varphi}\|_{4,q_1} \leq C \, \|\varphi\|_{3,p} \; . \tag{47}$$

Integrating (46) w.r.t. π_γ and using $\mathscr{L}\hat{\varphi} = \pi(\varphi) - \varphi$, we obtain that $\pi_\gamma(\varphi) - \pi(\varphi) = \gamma^{\alpha-1}\pi_\gamma(\mathscr{A}_\gamma \hat{\varphi})$. Combining this result with (47) concludes the proof.

6.3 Proof of Theorem 4

The proof is divided into two parts. In the first part, we derive some elementary bounds on the first and second order moments of the estimator $\hat{\pi}_n(f)$ defined in (1) and where $(X_k)_{k\in\mathbb{N}}$ is a Markov chain of kernel R_γ, see Lemma 12 below. The arguments are based solely on the study of R_γ and rely on $\mathbf{A}1(V, \bar{\gamma})$ and Lemma 2. We also provide an upper bound on $\mathbb{E}_{x,\gamma}\left[\{f(X_1) - R_\gamma f(x)\}^2\right]$ for $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, see Lemma 13. In a second part, we compare the discrete-time Markov chain $(X_k)_{k\in\mathbb{N}}$ and the Langevin diffusion $(Y_t)_{t\geq 0}$, see Lemma 14. The proof of Theorem 4 is then derived by a bootstrap argument based on Lemma 14. In the sequel, C is a non-negative constant independent of $\gamma > 0$ which may take different values at each appearance.

Note that if (25) holds, we obtain by Jensen's inequality and using $\sqrt{a+b} - \sqrt{a} \leq b/(2\sqrt{a})$ for all $a, b > 0$,

$$R_\gamma V^{1/2} \leq (\lambda^\gamma V + \gamma b)^{1/2} \leq \lambda^{\gamma/2} V^{1/2} + \gamma b \lambda^{-\gamma/2}/2 \; . \tag{48}$$

Lemma 12. *Let $V : \mathbb{R}^d \to [1, +\infty)$ and $\bar{\gamma} > 0$. Assume $\mathbf{A}1(V, \bar{\gamma})$. There exists $C > 0$ such that for all $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$, $n \in \mathbb{N}^*$ and $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$,*

$$\left| \mathbb{E}_{x,\gamma}\left[\sum_{k=0}^{n-1} \{f(X_k) - \pi_\gamma(f)\}\right] \right| \leq C\gamma^{-1} \, \|f\|_{V^{1/2}} \, V^{1/2}(x) \; , \tag{49}$$

$$\mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1} \{f(X_k) - \pi_\gamma(f)\}\right)^2\right] \leq C\gamma^{-2} \, \|f\|^2_{V^{1/2}} \left\{n + \gamma^{-1} V(x)\right\} \; . \tag{50}$$

Proof. By (48), $\mathbf{A}1(V^{1/2}, \bar{\gamma})$ is satisfied and using Lemma 2 with $V^{1/2}$, and $1 - \rho^\gamma \geq \gamma \ln(1/\rho)\rho^\gamma$, we obtain for all $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$,

$$\sum_{k=0}^{+\infty} \left| R^k_\gamma \{f - \pi_\gamma(f)\}(x) \right| \leq C\gamma^{-1} \, \|f\|_{V^{1/2}} \, V^{1/2}(x) \; , \tag{51}$$

which gives (49). For all $\gamma \in (0, \bar{\gamma}]$, define $\hat{f}_\gamma : \mathbb{R}^d \to \mathbb{R}$ by $\hat{f}_\gamma(x) = \sum_{k=0}^{+\infty} R_\gamma^k \{f - \pi_\gamma(f)\}(x)$, which is a solution of the Poisson equation, $(R_\gamma - \mathrm{Id})\hat{f}_\gamma = -\{f - \pi_\gamma(f)\}$, see [MT09, Section 17.4.1]. We get for all $n \in \mathbb{N}^*$,

$$\sum_{k=0}^{n-1} \{f(X_k) - \pi_\gamma(f)\} = \hat{f}_\gamma(X_0) - \hat{f}_\gamma(X_n) + \sum_{k=0}^{n-1} \{\hat{f}_\gamma(X_{k+1}) - R_\gamma \hat{f}_\gamma(X_k)\} \,. \tag{52}$$

By (51), for all $x \in \mathbb{R}^d$,

$$\hat{f}_\gamma^2(x) \leq C\gamma^{-2} \|f\|_{V^{1/2}}^2 V(x) \tag{53}$$

and $(\sum_{k=0}^{n-1} \{\hat{f}_\gamma(X_{k+1}) - R_\gamma \hat{f}_\gamma(X_k)\})_{n\in\mathbb{N}}$ is a square integrable martingale under $\mathbb{P}_{x,\gamma}$ for all $x \in \mathbb{R}^d$. By (52), we have

$$\begin{aligned} \mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1} \{f(X_k) - \pi_\gamma(f)\}\right)^2\right] &\leq 2\mathbb{E}_{x,\gamma}\left[\left(\hat{f}_\gamma(X_0) - \hat{f}_\gamma(X_n)\right)^2\right] \\ &\quad + 2\mathbb{E}_{x,\gamma}\left[\sum_{k=0}^{n-1} \{\hat{f}_\gamma(X_{k+1}) - R_\gamma \hat{f}_\gamma(X_k)\}^2\right] \,. \end{aligned} \tag{54}$$

Set $g_\gamma(x) = \mathbb{E}_{x,\gamma}\left[\{\hat{f}_\gamma(X_1) - R_\gamma \hat{f}_\gamma(x)\}^2\right]$. By (29) and (53), for all $x \in \mathbb{R}^d$, $g_\gamma(x) \leq C\gamma^{-2} \|f\|_{V^{1/2}}^2 V(x)$ and $\pi_\gamma(g_\gamma) \leq C\gamma^{-2} \|f\|_{V^{1/2}}^2$. By (51) for $f = g_\gamma$, we obtain

$$\left| \sum_{k=0}^{n-1} \{\mathbb{E}_{x,\gamma}[g_\gamma(X_k)] - \pi_\gamma(g_\gamma)\} \right| \leq C\gamma^{-3} \|f\|_{V^{1/2}}^2 V(x) \,.$$

Combining this result with (54), we obtain (50). □

Lemma 13. *Let $\bar{\gamma} > 0$ and $k \in \mathbb{N}$. Assume that $U \in \mathrm{C}_{\mathrm{poly}}^{k+1}(\mathbb{R}^d, \mathbb{R})$ and* **A** *2$(\bar{\gamma}, k)$. For any $p \in \mathbb{N}$, there exist $q \in \mathbb{N}$ and $C \geq 0$ such that for all $\gamma \in (0, \bar{\gamma}]$ and $f \in \mathrm{C}_{\mathrm{poly}}^{k+4}(\mathbb{R}^d, \mathbb{R})$, $\|\tilde{f}_\gamma\|_{k,q} \leq C\gamma \|f\|_{k+4,p}^2$ where $\tilde{f}_\gamma : \mathbb{R}^d \to \mathbb{R}$ is defined for all $x \in \mathbb{R}^d$ by*

$$\tilde{f}_\gamma(x) = \mathbb{E}_{x,\gamma}\left[\{f(X_1) - R_\gamma f(x)\}^2\right] \,.$$

Proof. Let $p \in \mathbb{N}$. By **A**2$(\bar{\gamma}, k)$, for all $f \in \mathrm{C}_{\mathrm{poly}}^{k+4}(\mathbb{R}^d, \mathbb{R})$, $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$,

$$\begin{aligned} \tilde{f}_\gamma(x) &= \mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x) - \gamma\mathscr{L}f(x) - \gamma^\alpha \mathscr{A}_\gamma f(x)\}^2\right] \\ &= \mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x)\}^2\right] + \gamma^2 \left\{\mathscr{L}f(x) + \gamma^{\alpha-1}\mathscr{A}_\gamma f(x)\right\}^2 \\ &\quad - 2\gamma \left\{\mathscr{L}f(x) + \gamma^{\alpha-1}\mathscr{A}_\gamma f(x)\right\} \mathbb{E}_{x,\gamma}[f(X_1) - f(x)] \,. \end{aligned} \tag{55}$$

Besides, for all $\gamma \in (0, \bar{\gamma}]$ and $x \in \mathbb{R}^d$,

$$
\begin{aligned}
\mathbb{E}_{x,\gamma}\left[\{f(X_1) - f(x)\}^2\right] &= \mathbb{E}_{x,\gamma}\left[f^2(X_1)\right] + f^2(x) - 2f(x)\mathbb{E}_{x,\gamma}\left[f(X_1)\right] \\
&= \gamma\mathscr{L}(f^2)(x) + \gamma^\alpha\mathscr{A}_\gamma(f^2)(x) - 2\gamma f(x)\mathscr{L}f(x) - 2\gamma^\alpha f(x)\mathscr{A}_\gamma f(x) \\
&= \gamma\left\{2\left\|\nabla f(x)\right\|^2 + \gamma^{\alpha-1}\left(\mathscr{A}_\gamma(f^2)(x) - 2f(x)\mathscr{A}_\gamma f(x)\right)\right\} \qquad (56)
\end{aligned}
$$

and $\mathbb{E}_{x,\gamma}\left[f(X_1) - f(x)\right] = \gamma\mathscr{L}f(x) + \gamma^\alpha\mathscr{A}_\gamma f(x)$. Then, combining (55) and (56), under **A** 2$(\bar{\gamma}, k)$, $\tilde{f}_\gamma \in \mathrm{C}^k_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and there exist $q \in \mathbb{N}$ and $C \geq 0$ such that $\|\tilde{f}_\gamma\|_{k,q} \leq C\gamma\|f\|^2_{k+4,p}$. □

Lemma 14. *Let $V : \mathbb{R}^d \to [1, +\infty)$ and $\bar{\gamma} > 0$. Assume **H** 1(4), **A** 1$(V, \bar{\gamma})$ and **A** 2$(\bar{\gamma}, 0)$. Then, for all $p \in \mathbb{N}$, there exists $C > 0$ such that for all $f \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}^*$,*

$$
\begin{aligned}
&\left|\mathbb{E}_{x,\gamma}\left[\frac{1}{n}\left(\sum_{k=0}^{n-1}\{f(X_k) - \pi_\gamma(f)\}\right)^2\right] - \frac{\sigma^2_\infty(f)}{\gamma}\right| \leq C\left\{\|f\|^2_{3,p}\,\gamma^{(\alpha-2)\wedge 0} + \frac{\|f\|^2_{3,p}V(x)}{n\gamma^2}\right. \\
&\left.+A_1^f(x,n,\gamma) + \frac{\|f\|^2_{3,p}V^{1/2}(x)\gamma^{(\alpha/2-1)\wedge 0}}{n^{1/2}\gamma} + A_1^f(x,n,\gamma)^{1/2}\,\|f\|_{3,p}\left(\gamma^{-1/2} + \frac{V^{1/2}(x)}{n^{1/2}\gamma}\right)\right\}, \qquad (57)
\end{aligned}
$$

where $\sigma^2_\infty(f)$ is defined in (7) and

$$
A_1^f(x,n,\gamma) = \frac{\gamma^{2(\alpha-1)}}{n}\mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1}\left\{\mathscr{A}_\gamma\hat{f}(X_k) - \gamma^{1-\alpha}\left(\pi_\gamma(f) - \pi(f)\right)\right\}\right)^2\right], \qquad (58)
$$

*$\alpha, \mathscr{A}_\gamma$ are given in **A** 2$(\bar{\gamma}, 0)$, and $\hat{f}$ is a solution of $\mathscr{L}\hat{f} = -\{f - \pi(f)\}$. Moreover,*

$$
A_1^f(x,n,\gamma) \leq C\gamma^{2(\alpha-2)}\,\|f\|^2_{3,p}\left\{1 + V(x)/(n\gamma)\right\}. \qquad (59)
$$

Proof. Under **H** 1(4) and by Proposition 1-(iii), let $\hat{f} \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ be a solution of the Poisson equation $\mathscr{L}\hat{f} = -\{f - \pi(f)\}$. Under **A** 2$(\bar{\gamma}, 0)$, we get for all $\gamma \in (0, \bar{\gamma}]$,

$$
R_\gamma\hat{f} = \hat{f} + \gamma\mathscr{L}\hat{f} + \gamma^\alpha\mathscr{A}_\gamma\hat{f}. \qquad (60)
$$

Since $\mathscr{L}\hat{f} = -\{f - \pi(f)\}$, we have for all $n \in \mathbb{N}^*$ and $\gamma \in (0, \bar{\gamma}]$,

$$
\begin{aligned}
\sum_{k=0}^{n-1}\{f(X_k) - \pi_\gamma(f)\} &= \frac{\hat{f}(X_0) - \hat{f}(X_n)}{\gamma} + \frac{1}{\gamma}\sum_{k=0}^{n-1}\left\{\hat{f}(X_{k+1}) - R_\gamma\hat{f}(X_k)\right\} \\
&\quad + \gamma^{\alpha-1}\sum_{k=0}^{n-1}\left\{\mathscr{A}_\gamma\hat{f}(X_k) - \gamma^{1-\alpha}\left(\pi_\gamma(f) - \pi(f)\right)\right\}. \qquad (61)
\end{aligned}
$$

Consider the following decomposition

$$\mathbb{E}_{x,\gamma}\left[\frac{1}{n}\left(\sum_{k=0}^{n-1}\{f(X_k)-\pi_\gamma(f)\}\right)^2\right]=\sum_{i=1}^{4}A_i^f(x,n,\gamma)\;, \tag{62}$$

where $A_1^f(x,n,\gamma)$ is given in (58),

$$A_2^f(x,n,\gamma)=\mathbb{E}_{x,\gamma}\left[(\hat{f}(X_0)-\hat{f}(X_n))^2/(n\gamma^2)\right]\;, \tag{63}$$

$$A_3^f(x,n,\gamma)=\mathbb{E}_{x,\gamma}\left[\frac{1}{n\gamma^2}\left(\sum_{k=0}^{n-1}\hat{f}(X_{k+1})-R_\gamma\hat{f}(X_k)\right)^2\right]\;,$$

and by Cauchy-Schwarz inequality,

$$(1/2)\left|A_4^f(x,n,\gamma)\right|\leq\sum_{1\leq i<j\leq 3}A_i^f(x,n,\gamma)^{1/2}A_j^f(x,n,\gamma)^{1/2}\;. \tag{64}$$

We show below that $\max_{i\in\{1,\ldots,4\}}\left|A_i^f(x,n,\gamma)\right|<+\infty$ for any $f\in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$. By Proposition 1-(iii), there exists $q_1\in\mathbb{N}$ such that $\|\hat{f}\|_{4,q_1}\leq C\|f\|_{3,p}$ and combining it with $\mathbf{A}1(V,\bar{\gamma})$ and (29), we obtain for all $x\in\mathbb{R}^d$, $\gamma\in(0,\bar{\gamma}]$ and $n\in\mathbb{N}^*$,

$$A_2^f(x,n,\gamma)\leq C\,\|f\|_{3,p}^2\,V(x)/(n\gamma^2)\;. \tag{65}$$

For all $\gamma\in(0,\bar{\gamma}]$ and $x\in\mathbb{R}^d$, set $g_\gamma(x)=\mathbb{E}_{x,\gamma}\left[\left\{\hat{f}(X_1)-R_\gamma\hat{f}(x)\right\}^2\right]$. By Proposition 1-(iii) and Lemma 13 with $k=0$, $g_\gamma\in\mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and for all $\gamma\in(0,\bar{\gamma}]$, $\|g_\gamma\|_V\leq C\gamma\|f\|_{3,p}^2$. Since $(\sum_{k=0}^{n-1}\hat{f}(X_{k+1})-R_\gamma\hat{f}(X_k))_{k\in\mathbb{N}}$ is a $\mathbb{P}_{x,\gamma}$-square integrable martingale, for all $x\in\mathbb{R}^d$, $n\in\mathbb{N}^*$ and $\gamma\in(0,\bar{\gamma}]$, we have by the Markov property

$$A_3^f(x,n,\gamma)=\gamma^{-2}\mathbb{E}_{x,\gamma}\left[\frac{1}{n}\sum_{k=0}^{n-1}g_\gamma(X_k)\right]\;. \tag{66}$$

By Lemma 12, eq. (49), we have for all $x\in\mathbb{R}^d$, $\gamma\in(0,\bar{\gamma}]$ and $n\in\mathbb{N}^*$,

$$\left|\mathbb{E}_{x,\gamma}\left[\frac{1}{n}\sum_{k=0}^{n-1}g_\gamma(X_k)\right]-\pi_\gamma(g_\gamma)\right|\leq C\,\|g_\gamma\|_V\,\frac{V(x)}{n\gamma}\leq C\,\|f\|_{3,p}^2\,\frac{V(x)}{n}\;. \tag{67}$$

We now show that $\pi_\gamma(g_\gamma)$ is approximately equal to $\gamma\sigma_\infty^2(f)$. Observe that

$$\begin{aligned}\pi_\gamma(g_\gamma)&=\mathbb{E}_{\pi_\gamma,\gamma}\left[\left\{\hat{f}(X_1)-R_\gamma\hat{f}(X_0)\right\}^2\right]\\&=\mathbb{E}_{\pi_\gamma,\gamma}\left[\left\{\hat{f}(X_1)-\hat{f}(X_0)\right\}^2\right]-\mathbb{E}_{\pi_\gamma,\gamma}\left[\left\{\hat{f}(X_0)-R_\gamma\hat{f}(X_0)\right\}^2\right]\;.\end{aligned} \tag{68}$$

We have by (60)

$$\mathbb{E}_{\pi_\gamma,\gamma}\left[\left\{\hat{f}(X_1)-\hat{f}(X_0)\right\}^2\right] = 2\mathbb{E}_{\pi_\gamma,\gamma}\left[\hat{f}(X_0)\left\{\hat{f}(X_0)-R_\gamma\hat{f}(X_0)\right\}\right] \tag{69}$$

$$= -2\gamma\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) - 2\gamma^\alpha\pi_\gamma(\hat{f}\mathscr{A}_\gamma\hat{f}) \; . \tag{70}$$

In the next step, we consider separately the cases $\pi_\gamma = \pi$ and $\pi_\gamma \neq \pi$.

- If $\pi = \pi_\gamma$, $-\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) = (1/2)\sigma^2_\infty(f)$.
- If $\pi_\gamma \neq \pi$, $(-\mathscr{L}\hat{f})\hat{f} \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and by Proposition 3, for all $\gamma \in (0,\bar{\gamma}]$,

$$\left|\pi_\gamma(\hat{f}\mathscr{L}\hat{f}) - \pi(\hat{f}\mathscr{L}\hat{f})\right| \leq C\left\|f\right\|^2_{3,p}\gamma^{\alpha-1} \; .$$

In both cases, using $\mathbf{A}2(\bar{\gamma},0)$, (29) and $\left|\pi_\gamma(\hat{f}\mathscr{A}_\gamma\hat{f})\right| \leq C\|f\|^2_{3,p}$, (70) becomes

$$\left|\mathbb{E}_{\pi_\gamma,\gamma}\left[\left\{\hat{f}(X_1)-\hat{f}(X_0)\right\}^2\right] - \gamma\sigma^2_\infty(f)\right| \leq C\left\|f\right\|^2_{3,p}\gamma^\alpha \; . \tag{71}$$

By $\mathbf{A}2(\bar{\gamma},0)$, (29) and (60), $\mathbb{E}_{\pi_\gamma,\gamma}\left[\left\{\hat{f}(X_0)-R_\gamma\hat{f}(X_0)\right\}^2\right] \leq C\|f\|^2_{3,p}\gamma^2$. Combining this result, (68) and (71),

$$\left|\pi_\gamma(g_\gamma) - \gamma\sigma^2_\infty(f)\right| \leq C\left\|f\right\|^2_{3,p}\gamma^{\alpha\wedge 2} \; . \tag{72}$$

Combining it with (67), for all $x \in \mathbb{R}^d$, $\gamma \in (0,\bar{\gamma}]$ and $n \in \mathbb{N}^*$,

$$\left|A^f_3(x,n,\gamma) - \frac{\sigma^2_\infty(f)}{\gamma}\right| \leq C\left\|f\right\|^2_{3,p}\left\{\gamma^{(\alpha-2)\wedge 0} + \frac{V(x)}{n\gamma^2}\right\} \; . \tag{73}$$

Combining (62), (64), (65) and (73) give (57).

For any $\gamma \in (0,\bar{\gamma}]$, by (60), $\pi_\gamma(\mathscr{A}_\gamma\hat{f}) = \gamma^{1-\alpha}\{\pi_\gamma(f)-\pi(f)\}$. Hence, by Lemma 12 and $\mathbf{A}2(\bar{\gamma},0)$, there exists $q_3 \in \mathbb{N}$ such that for all $x \in \mathbb{R}^d$ and $n \in \mathbb{N}^*$,

$$\begin{aligned} A^f_1(x,n,\gamma) &\leq C\gamma^{2(\alpha-2)}\left\|\mathscr{A}_\gamma\hat{f}\right\|^2_{V^{1/2}}\{1+V(x)/(n\gamma)\} \\ &\leq C\gamma^{2(\alpha-2)}\left\|\mathscr{A}_\gamma\hat{f}\right\|^2_{0,q_3}\{1+V(x)/(n\gamma)\} \\ &\leq C\gamma^{2(\alpha-2)}\left\|f\right\|^2_{3,p}\{1+V(x)/(n\gamma)\} \; , \end{aligned} \tag{74}$$

which gives (59). Finally, for any $f \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, there exists $p_f \in \mathbb{N}$ such that $\|f\|_{3,p_f} < +\infty$, and by (64), (65), (73) and (74), $\max_{i\in\{1,\ldots,4\}}\left|A^f_i(x,n,\gamma)\right| < +\infty$. □

of Theorem 4. To get the result, we use a bootstrap argument based on Lemma 14. Let $\hat{f} \in \mathrm{C}^7_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ be given by Proposition 1-(iii). We first apply Lemma 14 to the function $\mathscr{A}_\gamma \hat{f} \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. Note that by Proposition 1-(iii), $\mathbf{A}2(\bar{\gamma}, 3)$ and (59), there exist $q_1, q_2 \in \mathbb{N}$ such that for all $\gamma \in (0, \bar{\gamma}]$, $x \in \mathbb{R}^d$ and $n \in \mathbb{N}^*$,

$$\left\| \mathscr{A}_\gamma \hat{f} \right\|_{3,q_1} \leq C \left\| \hat{f} \right\|_{7,q_2} \leq C \left\| f \right\|_{6,p} ,$$
$$A_1^{\mathscr{A}_\gamma \hat{f}}(x, n, \gamma) \leq C \gamma^{2(\alpha-2)} \left\| f \right\|^2_{6,p} \{1 + V(x)/(n\gamma)\} .$$

By Lemma 14 applied to the function $\mathscr{A}_\gamma \hat{f}$, (57) and using $\pi_\gamma(\mathscr{A}_\gamma \hat{f}) = \gamma^{1-\alpha}\{\pi_\gamma(f) - \pi(f)\}$, we obtain for all $x \in \mathbb{R}^d$, $\gamma \in (0, \bar{\gamma}]$ and $n \in \mathbb{N}^*$

$$\begin{aligned} A_1^f(x, n, \gamma) &= \frac{\gamma^{2(\alpha-1)}}{n} \mathbb{E}_{x,\gamma}\left[\left(\sum_{k=0}^{n-1} \left\{ \mathscr{A}_\gamma \hat{f}(X_k) - \gamma^{1-\alpha} \left(\pi_\gamma(f) - \pi(f)\right) \right\} \right)^2 \right] \\ &\leq C \left\| f \right\|^2_{6,p} \gamma^{2(\alpha-1)} \left\{ \gamma^{-1} + \frac{V^{1/2}(x) \gamma^{(\alpha/2-1)\wedge 0}}{n^{1/2}\gamma} + \frac{V(x)}{n\gamma^2} \right\} . \end{aligned} \tag{75}$$

Combining Lemma 14 applied to the function $f \in \mathrm{C}^6_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, (57) and the upper bound (75) give

$$\begin{aligned} &\left| \mathbb{E}_{x,\gamma}\left[\frac{1}{n} \left(\sum_{k=0}^{n-1} \{f(X_k) - \pi_\gamma(f)\} \right)^2 \right] - \frac{\sigma^2_\infty(f)}{\gamma} \right| \leq C \left\| f \right\|^2_{6,p} \left\{ \gamma^{(\alpha-2)\wedge 0} \right. \\ &\left. + \frac{V^{1/2}(x)\gamma^{(\alpha/2-1)\wedge 0}}{n^{1/2}\gamma} + \frac{V^{1/4}(x)\gamma^{(\alpha/4-1/2)\wedge 0}}{n^{1/4}\gamma^{3/2-\alpha}} \left(\gamma^{-1/2} + \frac{V^{1/2}(x)}{n^{1/2}\gamma} \right) + \frac{V(x)}{n\gamma^2} \right\} . \end{aligned} \tag{76}$$

Note that by Young's inequality, we get for all $x \in \mathbb{R}^d$, $\gamma \in (0, \bar{\gamma}]$ and $n \in \mathbb{N}^*$,

$$\begin{aligned} \frac{V^{1/4}(x)}{n^{1/4}\gamma^{1/2}} \gamma^{(5\alpha/4-2)\wedge(\alpha-3/2)} &\leq \frac{1}{4} \frac{V(x)}{n\gamma^2} + \frac{3}{4} \gamma^{(5\alpha/3-8/3)\wedge(4\alpha/3-2)} , \\ 2\frac{V^{1/2}(x)\gamma^{(\alpha/2-1)\wedge 0}}{n^{1/2}\gamma} &\leq \frac{V(x)}{n\gamma^2} + \gamma^{(\alpha-2)\wedge 0} , \\ \frac{V(x)^{3/4}}{n^{3/4}\gamma^{3/2}} \gamma^{(5\alpha/4-3/2)\wedge(\alpha-1)} &\leq \frac{4}{3} \frac{V(x)}{n\gamma^2} + \frac{1}{4} \gamma^{(5\alpha-6)\wedge 4(\alpha-1)} . \end{aligned}$$

Combining it with the fact that for all $\alpha \geq 3/2$ and $\gamma \in (0, \bar{\gamma}]$,

$$\gamma^{(5\alpha/3-8/3)\wedge(4\alpha-2)} + \gamma^{(5\alpha-6)\wedge 4(\alpha-1)} \leq C \gamma^{(\alpha-2)\wedge 0} ,$$

concludes the proof. $\square$

Acknowledgements

This work was supported by the École Polytechnique Data Science Initiative.

References

[AC99] Roland Assaraf and Michel Caffarel. "Zero-variance principle for Monte Carlo algorithms". In: *Physical review letters* 83.23 (1999), p. 4682.

[Bak+08] D. Bakry et al. "A simple proof of the Poincaré inequality for a large class of probability measures." eng. In: *Electronic Communications in Probability [electronic only]* 13 (2008), pp. 60–66.

[BDM18] Nicolas Brosse, Alain Durmus, and Éric Moulines. "A scaling limit for the Random Walk Metropolis algorithm". In preparation. 2018.

[BGL14] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014, pp. xx+552.

[Bha82] R. N. Bhattacharya. "On Classical Limit Theorems for Diffusions". In: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 44.1 (1982), pp. 47–71.

[Bra+14] Silouanos Brazitikos et al. *Geometry of isotropic convex bodies*. Vol. 196. American Mathematical Society Providence, 2014.

[Bro+17] N. Brosse et al. "The Tamed Unadjusted Langevin Algorithm". In: *ArXiv e-prints* (Oct. 2017). arXiv: 1710.05559 [stat.ME].

[Bro+18] Nicolas Brosse et al. "Supplementary document 1: "Diffusion approximations and control variates for MCMC"". working paper or preprint. 2018.

[BS78] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control*. Vol. 139. Mathematics in Science and Engineering. The discrete time case. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1978, pp. xiii+323.

[CCG12] Patrick Cattiaux, Djalil Chafaı, and Arnaud Guillin. "Central limit theorems for additive functionals of ergodic Markov diffusions processes". In: *ALEA* 9.2 (2012), pp. 337–382.

[DK12] P. Dellaportas and I. Kontoyiannis. "Control variates for estimation based on reversible Markov chain Monte Carlo samplers". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74.1 (2012).

[DM16] A. Durmus and E. Moulines. "High-dimensional Bayesian inference via the Unadjusted Langevin Algorithm". In: *ArXiv e-prints* (May 2016). arXiv: 1605.01559 [math.ST].

[DM17] Alain Durmus and Éric Moulines. “Nonasymptotic convergence analysis for the unadjusted Langevin algorithm”. In: *Ann. Appl. Probab.* 27.3 (June 2017), pp. 1551–1587.

[Dou+18] Randal Douc et al. *Markov Chains.* Ed. by Springer. to be published. Springer, 2018.

[FHS15] Max Fathi, Ahmed-Amine Homman, and Gabriel Stoltz. “Error analysis of the transport properties of Metropolized schemes”. In: *ESAIM: Proc.* 48 (2015), pp. 341–363.

[FJ10] James M. Flegal and Galin L. Jones. “Batch means and spectral variance estimators in Markov chain Monte Carlo”. In: *Ann. Statist.* 38.2 (Apr. 2010), pp. 1034–1070.

[Gel+14] Andrew Gelman et al. *Bayesian data analysis.* Vol. 2. CRC press Boca Raton, FL, 2014.

[GM94] U. Grenander and M. I. Miller. “Representations of knowledge in complex systems”. In: *J. Roy. Statist. Soc. Ser. B* 56.4 (1994). With discussion and a reply by the authors, pp. 549–603.

[GM96] Peter W. Glynn and Sean P. Meyn. “A Liapunov bound for solutions of the Poisson equation”. In: *Ann. Probab.* 24.2 (Apr. 1996), pp. 916–931.

[GR14] Izrail Solomonovich Gradshteyn and Iosif Moiseevich Ryzhik. *Table of integrals, series, and products.* Academic press, 2014.

[Gre83] U. Grenander. “Tutorial in pattern theory”. Division of Applied Mathematics, Brown University, Providence. 1983.

[Hen97] Shane G Henderson. “Variance reduction via an approximating markov process”. Available at http://people.orie.cornell. edu/shane/pubs/thesis.pdf. PhD thesis. Department of Operations Research, Stanford University, 1997.

[HH14] Peter Hall and Christopher C Heyde. *Martingale limit theory and its application.* Academic press, 2014.

[Liu08] Jun S Liu. *Monte Carlo strategies in scientific computing.* Springer Science & Business Media, 2008.

[LS16] Tony Lelièvre and Gabriel Stoltz. “Partial differential equations and stochastic methods in molecular dynamics”. In: *Acta Numerica* 25 (2016), pp. 681–880.

[Mey08] Sean Meyn. *Control techniques for complex networks.* Cambridge University Press, 2008.

[MR07] Jean-Michel Marin and Christian Robert. *Bayesian core: a practical approach to computational Bayesian statistics.* Springer Science & Business Media, 2007.

[MSI13] Antonietta Mira, Reza Solgi, and Daniele Imparato. "Zero variance Markov chain Monte Carlo for Bayesian estimators". In: *Statistics and Computing* 23.5 (2013), pp. 653–662.

[MT09] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. 2nd. New York, NY, USA: Cambridge University Press, 2009.

[MV15] A. Mijatovic and J. Vogrinc. "On the Poisson equation for Metropolis-Hastings chains". In: *ArXiv e-prints* (Nov. 2015). arXiv: 1511.07464 [math.PR].

[MV17] A. Mijatović and J. Vogrinc. "Asymptotic variance for Random Walk Metropolis chains in high dimensions: logarithmic growth via the Poisson equation". In: *ArXiv e-prints* (July 2017). arXiv: 1707.08510 [math.PR].

[Oat+18] C. J. Oates et al. "Convergence Rates for a Class of Estimators Based on Stein's Method". In: *Accepted in Bernoulli* (Mar. 2018). arXiv: 1603.03220 [math.ST].

[OG16] Chris Oates and Mark Girolami. "Control Functionals for Quasi-Monte Carlo Integration". In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. Ed. by Arthur Gretton and Christian C. Robert. Vol. 51. Proceedings of Machine Learning Research. Cadiz, Spain: PMLR, May 2016, pp. 56–65.

[OGC16] Chris J. Oates, Mark Girolami, and Nicolas Chopin. "Control functionals for Monte Carlo integration". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2016), n/a–n/a.

[Par81] G. Parisi. "Correlation functions and computer simulations". In: *Nuclear Physics B* 180 (1981), pp. 378–384.

[PMG14] Theodore Papamarkou, Antonietta Mira, and Mark Girolami. "Zero Variance Differential Geometric Markov Chain Monte Carlo Algorithms". In: *Bayesian Anal.* 9.1 (Mar. 2014), pp. 97–128.

[PV01] E. Pardoux and Yu. Veretennikov. "On the Poisson Equation and Diffusion Approximation. I". In: *Ann. Probab.* 29.3 (July 2001), pp. 1061–1085.

[RC04] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Second. Springer Texts in Statistics. Springer-Verlag, New York, 2004, pp. xxx+645.

[RGG97] G. O. Roberts, A. Gelman, and W. R. Gilks. "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms". In: *The Annals of Applied Probability* 7.1 (1997), pp. 110–120.

[RK17] R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. Wiley Series in Probability and Statistics. Third edition [of MR0624270]. John Wiley & Sons, Inc., Hoboken, NJ, 2017, pp. xvii+414.

[RR98] Gareth O. Roberts and Jeffrey S. Rosenthal. "Optimal Scaling of Discrete Approximations to Langevin Diffusions". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 60.1 (1998), pp. 255–268.

[RT96] G. O. Roberts and R. L. Tweedie. “Exponential convergence of Langevin distributions and their discrete approximations”. In: *Bernoulli* 2.4 (1996), pp. 341–363.

A Strong Law of Large Numbers and Central Limit Theorem for the control variates estimator

Proposition 15. *Let $V : \mathbb{R}^d \to [1, +\infty)$, $\bar{\gamma} > 0$. Assume **A** 1$(V, \bar{\gamma})$ and that π_γ admits a positive density w.r.t. the Lebesgue measure for all $\gamma \in (0, \bar{\gamma}]$. Let $f \in \mathrm{C}_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $\psi = (\psi_1, \ldots, \psi_p) : \mathbb{R}^d \to \mathbb{R}^p$, $p \in \mathbb{N}^*$ be a fixed sieve of functions such that $(1, \psi_1, \ldots, \psi_p)$ is linearly independent in $\mathrm{C}(\mathbb{R}^d, \mathbb{R})$ and for all $i \in \{1, \ldots, p\}$, $\psi_i \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$. Then, for any initial probability measure ξ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*

$$\lim_{n \to +\infty} \hat{\theta}^*_n(f) = \theta^*_\gamma(f) \ , \quad \mathbb{P}_{\xi,\gamma} - a.s \ , \tag{77}$$

where $\hat{\theta}^*_n(f)$ and $\theta^*_\gamma(f)$ are defined in (18) and (21) respectively. Moreover, the following CLT holds for $\pi_n^{\mathrm{CV}}(f)$ defined in (17),

$$\sqrt{n} \left\{ \pi_n^{\mathrm{CV}}(f) - \pi_\gamma(f + \mathscr{L}(\theta^*_\gamma(f)^{\mathrm{T}} \psi)) \right\} \overset{\mathbb{P}_{\xi,\gamma}-weakly}{\underset{n \to +\infty}{\Longrightarrow}} \mathcal{N}(0, \sigma^2_{\infty,\gamma}(f + \mathscr{L}(\theta^*_\gamma(f)^{\mathrm{T}} \psi))) \ , \tag{78}$$

where $\sigma^2_{\infty,\gamma}(f + \mathscr{L}(\theta^*_\gamma(f)^{\mathrm{T}} \psi))$ is defined in (14).

Proof. By [Dou+18, Proposition 5.2.14], $\hat{\pi}_n(\psi \{f - \hat{\pi}_n(f)\})$ and H_n converges $\mathbb{P}_{\xi,\gamma}$-almost surely to $\pi_\gamma(\{f - \pi_\gamma(f)\} \psi)$ and H_γ where H_γ is a symmetric positive definite matrix defined in (20), and we obtain (77). Denote $W_{\gamma,n} \in \mathbb{R}^{p+1}$ for $n \in \mathbb{N}^*$ and $\gamma \in (0, \bar{\gamma}]$ by

$$W_{\gamma,n} = \sqrt{n} \left(\hat{\pi}_n(f) - \pi_\gamma(f), \hat{\pi}_n(\mathscr{L}\psi) - \pi_\gamma(\mathscr{L}\psi) \right) \ .$$

By [Dou+18, Proposition 21.1.3 and Theorem 21.2.11], $((1, \theta)^{\mathrm{T}} W_{\gamma,n})_{n \in \mathbb{N}^*}$ converges $\mathbb{P}_{\xi,\gamma}$-weakly, for every $\theta \in \mathbb{R}^p$ and any initial probability measure ξ, to a one-dimensional Gaussian variable of mean 0 and variance $\sigma^2_{\infty,\gamma}(f + \mathscr{L}(\theta^{\mathrm{T}} \psi))$. By the Cramér-Wold theorem, $(W_{\gamma,n})_{n \in \mathbb{N}^*}$ converges $\mathbb{P}_{\xi,\gamma}$-weakly to a $(p+1)$-dimensional Gaussian vector W_γ for any initial probability measure ξ, of mean 0 and covariance matrix

$$\pi_\gamma \left((\hat{f}_\gamma, \widehat{\mathscr{L}\psi}_\gamma)(\hat{f}_\gamma, \widehat{\mathscr{L}\psi}_\gamma)^{\mathrm{T}} - (R_\gamma \hat{f}_\gamma, R_\gamma \widehat{\mathscr{L}\psi}_\gamma)(R_\gamma \hat{f}_\gamma, R_\gamma \widehat{\mathscr{L}\psi}_\gamma)^{\mathrm{T}} \right) \ ,$$

where $\hat{f}_\gamma$ and $\widehat{\mathscr{L}\psi}_\gamma$ are solutions of the Poisson equations

$$(R_\gamma - \mathrm{Id}) \hat{f}_\gamma = -\left(f - \pi_\gamma(f) \right) \quad , \quad (R_\gamma - \mathrm{Id}) \widehat{\mathscr{L}\psi}_\gamma = -\left(\mathscr{L}\psi - \pi_\gamma(\mathscr{L}\psi) \right) \ .$$

By Slutsky's theorem, $(\hat{\theta}^*_n(f), W_{\gamma,n})_{n \in \mathbb{N}^*}$ converges $\mathbb{P}_{\xi,\gamma}$-weakly to $(\theta^*_\gamma(f), W_\gamma)$ and we obtain (78). □

Note that for the MALA and RWM algorithms, $\pi_\gamma = \pi$ and π has a positive density w.r.t. Leb (where Leb denotes the Lebesgue measure on $\mathbb{R}^d$). For ULA, since R_γ^{ULA} is Leb-irreducible for $\gamma > 0$, Leb is absolutely continuous w.r.t. π_γ. Indeed, π_γ is a maximal irreducibility measure and then Leb $\ll \pi_\gamma$.

B Law of Large Numbers and Central Limit Theorem for a step size γ_n function of the number of samples n

In this Section, we move away from the formalism of the canonical space to construct iteratively an array of Markov chains on the same filtered probability space, which allows us to give a precise meaning to the convergence in law of Theorem 5. Note first that every homogeneous Markov chain $(X_k)_{k\in\mathbb{N}}$ with values in $\mathbb{R}^d$ can be represented as a random iterative sequence, *i.e.* $X_{k+1} = F(X_k, \zeta_{k+1})$, where $(\zeta_k)_{k\in\mathbb{N}^*}$ is an i.i.d. sequence of uniform random variables on $[0, 1]$, X_0 is independent of $(\zeta_k)_{k\in\mathbb{N}^*}$ and F is a measurable function. See for example [Dou+18, Section 1.3.2] for a proof for $\mathbb{R}$-valued Markov chains, which can be extended to any Polish space by Kuratowski's theorem [BS78, Corollary 7.16.1].

Let $(\zeta_k)_{k\in\mathbb{N}^*}$ be an i.i.d. sequence of uniform random variables on $[0, 1]$ and Ξ be a random variable distributed according to the initial probability measure ξ, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the filtration $(\mathcal{F}_k)_{k\in\mathbb{N}}$ defined for all $k \in \mathbb{N}$ by $\mathcal{F}_k = \sigma(\Xi, \zeta_1, \ldots, \zeta_k)$. Let $(\gamma_n)_{n\in\mathbb{N}^*}$ be a positive sequence. By the preceding discussion, for all $n \in \mathbb{N}^*$, there exists a Borel measurable function $F_{\gamma_n} : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d$ such that the process $(X_k^n)_{k\in\mathbb{N}}$ defined for all $k \in \mathbb{N}$ by

$$X_{k+1}^n = F_{\gamma_n}(X_k^n, \zeta_{k+1}) \quad \text{and} \quad X_0^n = \Xi \,, \tag{79}$$

is a Markov chain on $(\Omega, (\mathcal{F}_k)_{k\in\mathbb{N}})$ associated with the Markov kernel R_{γ_n}.

In the sequel, C is a non-negative constant independent of $n \in \mathbb{N}^*$ which may take different values at each appearance. We first derive a Law of Large Numbers for the array $\{(X_k^n)_{k\in\{0,\ldots,n-1\}},\ n \in \mathbb{N}\}$ in Lemma 16. As an application, we show in Lemma 17 that $\hat{\theta}_n^*(f)$ converges in probability to $\theta^*(f)$ for a smooth $f : \mathbb{R}^d \to \mathbb{R}$, where $\hat{\theta}_n^*(f)$ and $\theta^*(f)$ are defined in (18) and (13), relatively to $(X_k^n)_{k\in\mathbb{N}}$. A Central Limit Theorem is provided in Proposition 18. Combining these results, we obtain the proof of Theorem 5.

Lemma 16. *Let $V : \mathbb{R}^d \to [1, +\infty)$, $\bar{\gamma} > 0$. Assume* ***H*** *1(4),* ***A*** *1$(V, \bar{\gamma})$, and* ***A*** *2$(\bar{\gamma}, 0)$. Let $\{(X_k^n)_{k\in\{0,\ldots,n-1\}},\ n \in \mathbb{N}\}$ be defined in (79) and assume that $\xi(V) < +\infty$. Let $f \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$, $(\gamma_n)_{n\in\mathbb{N}^*}$ be a positive sequence such that $\gamma_n \le \bar{\gamma}$ for all $n \in \mathbb{N}^*$, and $\lim_{n\to+\infty}(n\gamma_n)^{-1} + \gamma_n = 0$. Then,*

$$n^{-1}\sum_{k=0}^{n-1}\{f(X_k^n) - \pi(f)\} \underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}} 0 \,.$$

Proof. Let $f \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and $p \in \mathbb{N}$ such that $\|f\|_{3,p} < +\infty$. By Proposition 1-(iii), there exists $\hat{f} \in \mathrm{C}^4_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ such that $\mathscr{L}\hat{f} = -(f - \pi(f))$. By (61), we have for all $n \in \mathbb{N}^*$,

$$n^{-1}\sum_{k=0}^{n-1}\{f(X_k^n) - \pi(f)\} = \sum_{i=1}^{4} T_i^f(n) \,,$$

where

$$T_1^f(n) = (n\gamma_n)^{-1}\left\{\hat{f}(X_0^n) - \hat{f}(X_n^n)\right\} ,$$

$$T_2^f(n) = (n\gamma_n)^{-1}\sum_{k=0}^{n-1}\left\{\hat{f}(X_{k+1}^n) - R_{\gamma_n}\hat{f}(X_k^n)\right\} ,$$

$$T_3^f(n) = n^{-1}\gamma_n^{\alpha-1}\sum_{k=0}^{n-1}\mathscr{A}_{\gamma_n}\hat{f}(X_k^n) .$$

By (29), $\lim_{n\to+\infty}\mathbb{E}\left[\left(T_1^f(n)\right)^2\right] = 0$. Set $g_{\gamma_n}(x) = \mathbb{E}\left[\left(\hat{f}(X_1^n) - R_{\gamma_n}\hat{f}(x)\right)^2\right]$. By Lemma 13 with $k = 0$, $g_{\gamma_n} \in \mathrm{C}^0_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and there exists $q_1 \in \mathbb{N}$ such that for all $n \in \mathbb{N}^*$, $\|g_{\gamma_n}\|_{0,q_1} \leq C\gamma_n\|f\|^2_{3,p}$. By the Markov property and (29), we obtain for all $n \in \mathbb{N}^*$,

$$\mathbb{E}\left[\left(T_2^f(n)\right)^2\right] = (n\gamma_n)^{-2}\sum_{k=0}^{n-1}\mathbb{E}\left[g_{\gamma_n}(X_k^n)\right] \leq C(n\gamma_n)^{-2}n\left\|g_{\gamma_n}\right\|_{0,q_1} \leq C(n\gamma_n)^{-1}\left\|f\right\|^2_{3,p} ,$$

and $\lim_{n\to+\infty}\mathbb{E}\left[\left(T_2^f(n)\right)^2\right] = 0$ by assumption on $(\gamma_n)_{n\in\mathbb{N}^*}$. By **A** 2($\bar{\gamma}$, 0), there exists q_2 such that for all $n \in \mathbb{N}^*$, $\|\mathscr{A}_{\gamma_n}\hat{f}\|_{0,q_2} \leq C\|f\|_{3,p}$ and we get

$$\mathbb{E}\left[\left(T_3^f(n)\right)^2\right] \leq n^{-1}\gamma_n^{2(\alpha-1)}\sum_{k=0}^{n-1}\mathbb{E}\left[\left(\mathscr{A}_{\gamma_n}\hat{f}(X_k^n)\right)^2\right] \leq C\gamma_n^{2(\alpha-1)}\left\|f\right\|^2_{3,p} ,$$

and $\lim_{n\to+\infty}\mathbb{E}\left[\left(T_3^f(n)\right)^2\right] = 0$ by assumption on $(\gamma_n)_{n\in\mathbb{N}^*}$, which concludes the proof. □

Lemma 17. *Let $V : \mathbb{R}^d \to [1,+\infty)$, $\bar{\gamma} > 0$. Assume* ***H*** *1(4),* ***A*** *1($V,\bar{\gamma}$), and* ***A*** *2($\bar{\gamma}$, 0). Let $\{(X_k^n)_{k\in\{0,\dots,n-1\}},\ n \in \mathbb{N}\}$ be defined in (79) and assume that $\xi(V) < +\infty$. Let $f \in \mathrm{C}^3_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, $\psi = (\psi_1,\dots,\psi_p) : \mathbb{R}^d \to \mathbb{R}^p$, $p \in \mathbb{N}^*$ be a fixed sieve of functions such that $(1,\psi_1,\dots,\psi_p)$ is linearly independent in $\mathrm{C}(\mathbb{R}^d,\mathbb{R})$ and for all $i \in \{1,\dots,p\}$, $\psi_i \in \mathrm{C}^5_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$. Let $(\gamma_n)_{n\in\mathbb{N}^*}$ be a positive sequence such that $\gamma_n \leq \bar{\gamma}$ for all $n \in \mathbb{N}^*$, $\lim_{n\to+\infty}(n\gamma_n)^{-1} + \gamma_n = 0$. Then,*

$$\hat{\theta}_n^*(f) - \theta^*(f) \underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}} 0 ,$$

where $\hat{\theta}_n^*(f)$ and $\theta^*(f)$ are defined in (18) and (13), respectively, relatively to $(X_k^n)_{k\in\mathbb{N}}$.

Proof. For all $n \in \mathbb{N}^*$,

$$\hat{\theta}_n^*(f) - \theta^*(f) = \left(H_n^+ - H^{-1}\right)\hat{\pi}_n\left(\psi(f - \hat{\pi}_n(f))\right) + H^{-1}\left\{\hat{\pi}_n\left(\psi(f - \hat{\pi}_n(f))\right) - \pi\left(\psi(f - \pi(f))\right)\right\} .$$

By Lemma 16,

$$\hat{\pi}_n\left(\psi(f-\hat{\pi}_n(f))\right)-\pi\left(\psi(f-\pi(f))\right)\underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}}0\,,$$

and it is enough to show that

$$H_n^+-H^{-1}\underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}}0\,,$$

to conclude the proof. Let $\epsilon>0$ and consider the following decomposition:

$$\begin{aligned}\left\{\left\|H_n^+-H^{-1}\right\|\geq\epsilon\right\}=&\left\{\left\|H_n^+-H^{-1}\right\|\geq\epsilon\right\}\cap\left\{\left\|H^{-1}\right\|\left\|H_n-H\right\|\leq1/2\right\}\\&\cup\left\{\left\|H_n^+-H^{-1}\right\|\geq\epsilon\right\}\cap\left\{\left\|H^{-1}\right\|\left\|H_n-H\right\|>1/2\right\}\,,\end{aligned}$$

where $\|\cdot\|$ denotes the operator norm. Since by Lemma 16,

$$H_n-H\underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}}0\,,$$

we obtain

$$\begin{aligned}&\mathbb{P}\left(\left\{\left\|H_n^+-H^{-1}\right\|\geq\epsilon\right\}\cap\left\{\left\|H^{-1}\right\|\left\|H_n-H\right\|>1/2\right\}\right)\\&\qquad\leq\mathbb{P}\left(\left\|H^{-1}\right\|\left\|H_n-H\right\|>1/2\right)\underset{n\to+\infty}{\longrightarrow}0\,.\end{aligned}$$

By [Dou+18, Corollary 22.A.6], on the event $\left\{\left\|H^{-1}\right\|\left\|H_n-H\right\|\leq1/2\right\}$,

$$\left\|H_n^+-H^{-1}\right\|=\left\|H_n^{-1}-H^{-1}\right\|\leq\frac{\left\|H^{-1}\right\|^2\left\|H_n-H\right\|}{1-\left\|H^{-1}\right\|\left\|H_n-H\right\|}\leq2\left\|H^{-1}\right\|^2\left\|H_n-H\right\|\,,$$

and,

$$\begin{aligned}&\mathbb{P}\left(\left\{\left\|H_n^+-H^{-1}\right\|\geq\epsilon\right\}\cap\left\{\left\|H^{-1}\right\|\left\|H_n-H\right\|\leq1/2\right\}\right)\\&\qquad\leq\mathbb{P}\left(2\left\|H^{-1}\right\|^2\left\|H_n-H\right\|\geq\epsilon\right)\underset{n\to+\infty}{\longrightarrow}0\,,\end{aligned}$$

which gives the result. □

Proposition 18. *Let $V:\mathbb{R}^d\to[1,+\infty)$, $\bar{\gamma}>0$. Assume $\boldsymbol{H}$ 1(10), $\boldsymbol{A}$ 1($V,\bar{\gamma}$), and $\boldsymbol{A}$ 2($\bar{\gamma},6$). Let $\{(X_k^n)_{k\in\{0,\ldots,n-1\}},\ n\in\mathbb{N}\}$ be defined in (79) and assume that $\xi(V)<+\infty$. Let $f\in\mathrm{C}^9_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, $(\gamma_n)_{n\in\mathbb{N}^*}$ be a positive sequence satisfying $\lim_{n\to+\infty}(n\gamma_n)^{-1}+\gamma_n=0$ and $\hat{f}$ be a solution of the Poisson equation $\mathscr{L}\hat{f}=\pi(f)-f$. Then,*

(i) if $\pi(\mathscr{A}_0\hat{f})\lim_{n\to+\infty}n^{1/2}\gamma_n^{\alpha-1/2}=0$,

$$n^{-1/2}\gamma_n^{1/2}\sum_{k=0}^{n-1}\left\{f(X_k^n)-\pi(f)\right\}\underset{n\to+\infty}{\overset{\mathbb{P}-weakly}{\Longrightarrow}}\mathcal{N}(0,\sigma_\infty^2(f))\,,$$

(ii) if $\lim_{n\to+\infty} n^{1/2}\gamma_n^{\alpha-1/2} = \gamma_\infty \in [0,+\infty)$,

$$n^{-1/2}\gamma_n^{1/2}\sum_{k=0}^{n-1}\{f(X_k^n)-\pi(f)\} \underset{n\to+\infty}{\overset{\mathbb{P}-weakly}{\Longrightarrow}} \mathcal{N}(\gamma_\infty\pi(\mathscr{A}_0\hat{f}),\sigma_\infty^2(f)) ,$$

(iii) if $\pi(\mathscr{A}_0\hat{f})\liminf_{n\to+\infty} n^{1/2}\gamma_n^{\alpha-1/2} = +\infty$,

$$\gamma_n^{1-\alpha}\sum_{k=0}^{n-1}\{f(X_k^n)-\pi(f)\} \underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}} \pi(\mathscr{A}_0\hat{f}) \quad ,$$

where $\sigma_\infty^2(f)$ *is defined in* (7).

Note that if the invariant distribution of R_γ is π for all $\gamma \in (0,\bar{\gamma}]$ (e.g. the case of MALA or RWM), we have under $\mathbf{A}2(\bar{\gamma},0)$ and by the dominated convergence theorem, $\pi(\mathscr{A}_0\hat{f}) = 0$.

Proof. Let $f \in \mathrm{C}^9_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and $p\in\mathbb{N}$ such that $\|f\|_{9,p} < +\infty$. By Proposition 1-(iii), there exists $\hat{f} \in \mathrm{C}^{10}_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ such that $\mathscr{L}\hat{f} = -(f-\pi(f))$. By (61), we have for all $n\in\mathbb{N}^*$,

$$n^{-1/2}\gamma_n^{1/2}\sum_{k=0}^{n-1}\{f(X_k^n)-\pi(f)\} = \sum_{i=1}^{4} B_i^f(n) ,$$

where

$$\begin{aligned}
B_1^f(n) &= (n\gamma_n)^{-1/2}\left\{\hat{f}(X_0^n)-\hat{f}(X_n^n)\right\} ,\\
B_2^f(n) &= (n\gamma_n)^{-1/2}\sum_{k=0}^{n-1}\left\{\hat{f}(X_{k+1}^n)-R_{\gamma_n}\hat{f}(X_k^n)\right\} ,\\
B_3^f(n) &= n^{-1/2}\gamma_n^{1/2}\gamma_n^{\alpha-1}\sum_{k=0}^{n-1}\left\{\mathscr{A}_{\gamma_n}\hat{f}(X_k^n)-\pi_{\gamma_n}(\mathscr{A}_{\gamma_n}\hat{f})\right\} ,\\
B_4^f(n) &= n^{1/2}\gamma_n^{1/2}\gamma_n^{\alpha-1}\pi_{\gamma_n}(\mathscr{A}_{\gamma_n}\hat{f}) .
\end{aligned}$$

We show in the sequel that $B_1^f(n)$ and $B_3^f(n)$ are remainder terms that converge in probability to 0 as $n\to+\infty$. $B_2^f(n)$ converges in law to a Gaussian random variable of mean 0 and variance $\sigma_\infty^2(f)$. $B_4^f(n)$ is the bias term.

By (29), $\lim_{n\to+\infty}\mathbb{E}\left[\left|B_1^f(n)\right|\right] = 0$ and then $B_1^f(n)$ converges in probability to 0 as $n\to+\infty$. By $\mathbf{A}2(\bar{\gamma},6)$, $\mathscr{A}_{\gamma_n}\hat{f}\in\mathrm{C}^6_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and there exists $q_1\in\mathbb{N}$ such that for all $n\in\mathbb{N}^*$, $\|\mathscr{A}_{\gamma_n}\hat{f}\|_{6,q_1}\le C\|f\|_{9,p}$. We obtain by Theorem 4 and using that $\xi(V)<+\infty$,

$$\mathbb{E}\left[\left(B_3^f(n)\right)^2\right] \le C\left\|\mathscr{A}_{\gamma_n}\hat{f}\right\|_{6,p}^2\gamma_n^{2(\alpha-1)}\left\{1+\frac{1}{n\gamma_n}\right\} ,$$

and $\lim_{n\to+\infty} \mathbb{E}\left[\left(B_3^f(n)\right)^2\right] = 0$.

We now consider $B_2^f(n)$ for $n \in \mathbb{N}^*$. For $k \in \{0, \dots, n-1\}$, denote by

$$\theta_{k+1,n} = (n\gamma_n)^{-1/2} \left\{ \hat{f}(X_{k+1}^n) - R_{\gamma_n} \hat{f}(X_k^n) \right\} .$$

By [HH14, Corollary 3.1, Chapter 3], $B_2^f(n)$ converges in law to a Gaussian random variable of mean 0 and variance $\sigma_\infty^2(f)$ if

$$\sum_{k=0}^{n-1} \mathbb{E}\left[\theta_{k+1,n}^2 \middle| \mathcal{F}_k\right] \underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}} \sigma_\infty^2(f) , \tag{80}$$

$$\sum_{k=0}^{n-1} \mathbb{E}\left[\theta_{k+1,n}^4 \middle| \mathcal{F}_k\right] \underset{n\to+\infty}{\overset{\mathbb{P}}{\Longrightarrow}} 0 . \tag{81}$$

Set $g_{\gamma_n}(x) = \mathbb{E}\left[\left(\hat{f}(X_1^n) - R_{\gamma_n}\hat{f}(x)\right)^2\right]$. By Lemma 13 with $k = 6$, $g_{\gamma_n} \in \mathrm{C}^6_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ and there exists $q_2 \in \mathbb{N}$ such that for all $n \in \mathbb{N}^*$

$$\|g_{\gamma_n}\|_{6,q_2} \leq C\gamma_n \|f\|_{9,p}^2 . \tag{82}$$

By Proposition 1-(iii), for all $n \in \mathbb{N}^*$, there exists $\hat{g}_{\gamma_n} \in \mathrm{C}^7_{\mathrm{poly}}(\mathbb{R}^d, \mathbb{R})$ such that $\mathscr{L}\hat{g}_{\gamma_n} = -(g_{\gamma_n} - \pi_{\gamma_n}(g_{\gamma_n}))$. By the Markov property and (61), we have for all $n \in \mathbb{N}^*$

$$\sum_{k=0}^{n-1} \mathbb{E}\left[\theta_{k+1,n}^2 \middle| \mathcal{F}_k\right] = \frac{1}{n\gamma_n} \sum_{k=0}^{n-1} g_{\gamma_n}(X_k^n) = B_{21}^f(n) + B_{22}^f(n) + B_{23}^f(n) + B_{24}^f(n) ,$$

where

$$\begin{aligned}
B_{21}^f(n) &= \gamma_n^{-1} \pi_{\gamma_n}(g_{\gamma_n}) , \\
B_{22}^f(n) &= (n\gamma_n^2)^{-1} \left\{ \hat{g}_{\gamma_n}(X_0^n) - \hat{g}_{\gamma_n}(X_n^n) \right\} , \\
B_{23}^f(n) &= (n\gamma_n^2)^{-1} \sum_{k=0}^{n-1} \left\{ \hat{g}_{\gamma_n}(X_{k+1}^n) - R_{\gamma_n}\hat{g}_{\gamma_n}(X_k^n) \right\} , \\
B_{24}^f(n) &= \frac{1}{n\gamma_n} \gamma_n^{\alpha-1} \sum_{k=0}^{n-1} \left\{ \mathscr{A}_{\gamma_n}\hat{g}_{\gamma_n}(X_k^n) - \gamma_n^{1-\alpha} \left(\pi_{\gamma_n}(\hat{g}_{\gamma_n}) - \pi(\hat{g}_{\gamma_n}) \right) \right\} .
\end{aligned}$$

By (29),

$$\mathbb{E}\left[\left|B_{22}^f(n)\right|\right] \leq C(n\gamma_n^2)^{-1} \|g_{\gamma_n}\|_{6,q_2} \leq C(n\gamma_n)^{-1} \|f\|_{9,p}^2 ,$$

$\lim_{n\to+\infty} \mathbb{E}\left[\left|B_{22}^f(n)\right|\right] = 0$, and $B_{22}^f(n)$ converges in probability to 0 as $n \to +\infty$. By (29), (82), the Markov property and Lemma 13 with $k = 0$, we get for all $n \in \mathbb{N}^*$

$$\begin{aligned}
\mathbb{E}\left[\left(B_{23}^f(n)\right)^2\right] &= \frac{1}{n\gamma_n^4} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}\left[\left(\hat{g}_{\gamma_n}(X_{k+1}^n) - R_{\gamma_n}\hat{g}_{\gamma_n}(X_k^n)\right)^2\right] \\
&\leq C \frac{1}{n\gamma_n^4} \gamma_n \|g_{\gamma_n}\|_{6,q_2}^2 \leq C \frac{1}{n\gamma_n} \|f\|_{9,p}^4 ,
\end{aligned}$$

and $\lim_{n\to+\infty}\mathbb{E}\left[\left(B^f_{23}(n)\right)^2\right]=0$. We can decompose $B^f_{24}(n)$ as,

$$
\begin{aligned}
\mathbb{E}\left[\left(B^f_{24}(n)\right)^2\right] &= \frac{\gamma_n^{2(\alpha-1)}}{n\gamma_n^2}\mathbb{E}\left[\frac{1}{n}\left(\sum_{k=0}^{n-1}\left\{\mathscr{A}_{\gamma_n}\hat{g}_{\gamma_n}(X_k^n)-\gamma_n^{1-\alpha}\left(\pi_{\gamma_n}(\hat{g}_{\gamma_n})-\pi(\hat{g}_{\gamma_n})\right)\right\}\right)^2\right] \\
&= \frac{1}{n\gamma_n^3}B^f_{241}(n) \qquad (83)
\end{aligned}
$$

where,

$$
B^f_{241}(n)=\gamma_n^{2(\alpha-1)}\gamma_n\mathbb{E}\left[\frac{1}{n}\left(\sum_{k=0}^{n-1}\left\{\mathscr{A}_{\gamma_n}\hat{g}_{\gamma_n}(X_k^n)-\gamma_n^{1-\alpha}\left(\pi_{\gamma_n}(\hat{g}_{\gamma_n})-\pi(\hat{g}_{\gamma_n})\right)\right\}\right)^2\right].
$$

By **A** 2$(\bar{\gamma},6)$ and Proposition 1-(iii), there exists $q_3\in\mathbb{N}$ such that for all $n\in\mathbb{N}^*$, $\|\mathscr{A}_{\gamma_n}\hat{g}_{\gamma_n}\|_{3,q_3}\le C\|g_{\gamma_n}\|_{6,q_2}$. By Lemma 14, (59), (82) and using that $\xi(V)<+\infty$,

$$
\begin{aligned}
B^f_{241}(n) &\le C\gamma_n\gamma_n^{2(\alpha-2)}\left\|\mathscr{A}_{\gamma_n}\hat{g}_{\gamma_n}\right\|^2_{3,q_3}\{1+1/(n\gamma_n)\}\le C\left\|g_{\gamma_n}\right\|^2_{6,q_2}\{1+1/(n\gamma_n)\} \\
&\le C\gamma_n^2\left\|f\right\|^4_{9,p}\{1+1/(n\gamma_n)\}\ .
\end{aligned}
$$

Combining it with (83), we obtain $\lim_{n\to+\infty}\mathbb{E}\left[\left(B^f_{24}(n)\right)^2\right]=0$. For $B^f_{21}(n)$, we have by (72) and (82) for all $n\in\mathbb{N}^*$,

$$
\left|\gamma_n^{-1}\pi_{\gamma_n}(g_{\gamma_n})-\sigma^2_\infty(f)\right|\le C\left\|g_{\gamma_n}\right\|^2_{3,q_2}\gamma_n^{\alpha\wedge 2}\gamma_n^{-1}\le C\left\|f\right\|^4_{9,p}\gamma_n\gamma_n^{\alpha\wedge 2}\ ,
$$

and $\lim_{n\to+\infty}\gamma_n^{-1}\pi_{\gamma_n}(g_{\gamma_n})=\sigma^2_\infty(f)$. This gives (80). For (81), we have for $k\in\{0,\ldots,n-1\}$

$$
\mathbb{E}\left[\theta^4_{k+1,n}\middle|\mathcal{F}_k\right]=(n\gamma_n)^{-2}h_{\gamma_n}(X_k^n)
$$

where $h_{\gamma_n}(x)=\mathbb{E}\left[\left\{\hat{f}(X_1^n)-R_{\gamma_n}\hat{f}(x)\right\}^4\right]$. By Cauchy-Schwarz inequality and Lemma 13, there exists $q_4\in\mathbb{N}$ such that for all $n\in\mathbb{N}^*$, $\|h_{\gamma_n}\|_{0,q_4}\le C\gamma_n^2\|f\|^4_{3,p}$. By (29), we obtain for all $n\in\mathbb{N}^*$

$$
\mathbb{E}\left[\left|\sum_{k=0}^{n-1}\mathbb{E}\left[\theta^4_{k+1,n}\middle|\mathcal{F}_k\right]\right|\right]\le Cn^{-1}\left\|f\right\|^4_{3,p}
$$

and (81) is satisfied.

For $B^f_4(n)$, we only have to show that $\lim_{n\to+\infty}\pi_{\gamma_n}(\mathscr{A}_{\gamma_n}\hat{f})=\pi(\mathscr{A}_0\hat{f})$. Using Proposition 3, we have for all $n\in\mathbb{N}^*$,

$$
\left|\pi_{\gamma_n}(\mathscr{A}_{\gamma_n}\hat{f})-\pi(\mathscr{A}_{\gamma_n}\hat{f})\right|\le C\left\|\mathscr{A}_{\gamma_n}\hat{f}\right\|_{3,q_1}\gamma_n^{\alpha-1}\le C\left\|f\right\|_{9,p}\gamma_n^{\alpha-1}\ .
$$

Combining it with **A** 2$(\bar{\gamma},6)$ and the dominated convergence theorem, we obtain the result, which concludes the proof. □

of Theorem 5. We consider the case $\lim_{n\to+\infty} n^{1/2}\gamma_n^{\alpha-1/2} = \gamma_\infty \in [0,+\infty)$, and we denote by $\mu_f^{\mathrm{CV}} = \gamma_\infty \pi(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi) \in [0,+\infty)$. The case

$$\pi(\mathscr{A}_0(\hat{f} - \theta^*(f)^{\mathrm{T}}\psi)) \liminf_{n\to+\infty} n^{1/2}\gamma_n^{\alpha-1/2} = +\infty$$

can be handled in a similar way. Denote $W_n \in \mathbb{R}^{p+1}$ for $n \in \mathbb{N}^*$ by

$$W_n = n^{1/2}\gamma_n^{1/2}\left(\hat{\pi}_n(f) - \pi(f), \hat{\pi}_n(\mathscr{L}\psi)\right) \ .$$

By Proposition 18, $((1,\theta)^{\mathrm{T}} W_n)_{n\in\mathbb{N}^*}$ converges $\mathbb{P}$-weakly, for every $\theta \in \mathbb{R}^p$, to a one-dimensional Gaussian variable of mean μ_f^{CV} and variance $\sigma_\infty^2(f + \mathscr{L}(\theta^{\mathrm{T}}\psi))$. By the Cramér-Wold theorem, $(W_n)_{n\in\mathbb{N}^*}$ converges $\mathbb{P}$-weakly to a $(p+1)$-dimensional Gaussian vector W of mean $\gamma_\infty\left(\pi(\mathscr{A}_0\hat{f}), -\pi(\mathscr{A}_0\psi)\right)$ and covariance matrix

$$2\pi\left((\hat{f},-\psi)(-\mathscr{L})(\hat{f},-\psi)^{\mathrm{T}}\right) \ .$$

By Lemma 17 and Slutsky's theorem, $(\hat{\theta}_n^*(f), W_n)_{n\in\mathbb{N}^*}$ converges $\mathbb{P}$-weakly to $(\theta^*(f), W)$, which concludes the proof. □

C Additional proofs

C.1 Proof of Proposition 1

(i) By **H** 1(2), [RT96, Theorems 2.1], π is the unique stationary distribution of the semigroup $(P_t)_{t\geq 0}$ associated to (6). In addition, by [RT96, Theorems 2.1] and [Bak+08, Corollary 1.6], $(P_t)_{t\geq 0}$ is V-uniformly geometrically ergodic w.r.t. π with $V(x) = \exp\{(\upsilon/4)(1+\|x-x^\star\|^2)^{1/2}\}$.

(ii) is given by [GM96, Theorem 4.4] using that $(P_t)_{t\geq 0}$ is V-uniformly geometrically ergodic; see also see [Bha82] and [CCG12].

(iii) follows from [PV01, Theorem 1].

(iv) Let $f, g \in \mathrm{C}^2_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ and $M > 0$. We split $\pi(f(-\mathscr{L})g)$ into $I_1 + I_2$ where

$$I_1 = \int_{[-M,M]^d} f(x)(-\mathscr{L})g(x)\pi(\mathrm{d}x) \ , \quad I_2 = \int_{\left([-M,M]^d\right)^{\mathrm{c}}} f(x)(-\mathscr{L})g(x)\pi(\mathrm{d}x) \ .$$

By the dominated convergence theorem, $\lim_{M\to+\infty} I_2 = 0$. For all $i \in \{1,\ldots,d\}$, $a \in \mathbb{R}$ and $x \in \mathbb{R}^d$, denote by $x_{-i}^a = (x_1,\ldots,x_{i-1},a,x_{i+1},\ldots,x_d)$ and by $x_{-i} =$

$(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$. By integrations by parts,

$$
\begin{aligned}
I_1 &= \int_{[-M,M]^d} f(x) \langle \nabla U(x), \nabla g(x) \rangle \pi(\mathrm{d}x) + \int_{[-M,M]^d} f(x)(-\Delta g(x))\pi(\mathrm{d}x) \\
&= \int_{[-M,M]^d} f(x) \langle \nabla U(x), \nabla g(x) \rangle \pi(\mathrm{d}x) \\
&+ \sum_{i=1}^{d} \int_{[-M,M]^{d-1}} \left\{ f(x_{-i}^{-M}) \frac{\partial g}{\partial x_i}(x_{-i}^{-M}) \pi(x_{-i}^{-M}) - f(x_{-i}^{M}) \frac{\partial g}{\partial x_i}(x_{-i}^{M}) \pi(x_{-i}^{M}) \right\} \mathrm{d}x_{-i} \\
&+ \int_{[-M,M]^d} \{ \langle \nabla f(x), \nabla g(x) \rangle - f(x) \langle \nabla U(x), \nabla g(x) \rangle \} \pi(\mathrm{d}x)
\end{aligned}
$$

and $\lim_{M \to +\infty} I_1 = \pi(\langle \nabla f, \nabla g \rangle)$ which concludes the proof.

C.2 Proof of Lemma 6

By [DM16, Theorem 32], we have for $x, y \in \mathbb{R}^d$ and $\gamma > 0$

$$
\| \delta_x (R_\gamma^{\mathrm{ULA}})^{\lceil 1/\gamma \rceil} - \delta_y (R_\gamma^{\mathrm{ULA}})^{\lceil 1/\gamma \rceil} \|_{\mathrm{TV}} \le 1 - 2\Phi \left(- \frac{\|x - y\|}{2 \Xi_{\lceil 1/\gamma \rceil}^{1/2}} \right)
$$

where

$$
\Xi_{\lceil 1/\gamma \rceil} = \sum_{i=1}^{\lceil 1/\gamma \rceil} (2\gamma) \prod_{j=1}^{i} (1 + \gamma L)^{-2} = \frac{1 - \exp\left(-2 \lceil 1/\gamma \rceil \ln(1 + \gamma L)\right)}{L + (\gamma L^2)/2} ,
$$

which gives (i). The assertion (ii) follows from [DM17, Proposition 8] and [DM17, Proposition 13].

C.3 Proof of Lemma 8

We have for all $x \in \mathbb{R}^d$

$$
\begin{aligned}
\nabla U_{\log}(x) &= -\mathsf{X}^{\mathrm{T}} \mathsf{Y} + \sum_{i=1}^{N} \mathsf{X}_i / (1 + \mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}) + x/\varsigma^2 , \\
\mathrm{D}^2 U_{\log}(x) &= \sum_{i=1}^{N} \frac{\mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}}{\left(1 + \mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}\right)^2} \mathsf{X}_i \mathsf{X}_i^{\mathrm{T}} + \mathrm{Id} / \varsigma^2 , \\
\mathrm{D}^3 U_{\log}(x) &= \sum_{i=1}^{N} \frac{\mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}}{\left(1 + \mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}\right)^2} \left\{ 2 \frac{\mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}}{1 + \mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x}} - 1 \right\} \mathsf{X}_i^{\otimes 3} .
\end{aligned}
$$

Using for all $i \in \{1, \ldots, N\}$ and $x \in \mathbb{R}^d$ that $0 < \mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x} / (1 + \mathrm{e}^{-\mathsf{X}_i^{\mathrm{T}} x})^2 \le 1/4$, $U_{\log}$ is strongly convex, gradient Lipschitz and satisfies **H**2, **H**4, **H**1(k) for all $k \in \mathbb{N}^*$, **H**5 and **H**6.

For U_{pro}, define $h : \mathbb{R} \to \mathbb{R}_-$ for all $t \in \mathbb{R}$ by $h(t) = \ln(\Phi(t))$. We have for all $t \in \mathbb{R}$,

$$
h'(t) = \frac{\Phi'(t)}{\Phi(t)} \quad , \quad h''(t) = -\frac{\Phi'(t)}{\Phi(t)} \left\{ t + \frac{\Phi'(t)}{\Phi(t)} \right\} ,
$$

$$
h^{(3)}(t) = \frac{\Phi'(t)}{\Phi(t)} \left\{ 2 \left(\frac{\Phi'(t)}{\Phi(t)} \right)^2 + 3t \frac{\Phi'(t)}{\Phi(t)} + t^2 - 1 \right\}
$$

and for all $x \in \mathbb{R}^d$

$$
\nabla U_{\text{pro}}(x) = \sum_{i=1}^{N} \left\{ (1 - \mathsf{Y}_i) h'(-\mathsf{X}_i^{\mathrm{T}} x) - \mathsf{Y}_i h'(\mathsf{X}_i^{\mathrm{T}} x) \right\} \mathsf{X}_i + x/\varsigma^2 ,
$$

$$
\mathrm{D}^2 U_{\text{pro}}(x) = \sum_{i=1}^{N} \left\{ -(1 - \mathsf{Y}_i) h''(-\mathsf{X}_i^{\mathrm{T}} x) - \mathsf{Y}_i h''(\mathsf{X}_i^{\mathrm{T}} x) \right\} \mathsf{X}_i \mathsf{X}_i^{\mathrm{T}} + \operatorname{Id}/\varsigma^2 ,
$$

$$
\mathrm{D}^3 U_{\text{pro}}(x) = \sum_{i=1}^{N} \left\{ (1 - \mathsf{Y}_i) h^{(3)}(-\mathsf{X}_i^{\mathrm{T}} x) - \mathsf{Y}_i h^{(3)}(\mathsf{X}_i^{\mathrm{T}} x) \right\} \mathsf{X}_i^{\otimes 3} .
$$

By an integration by parts, we have for all $t < 0$

$$
t + \frac{\Phi'(t)}{\Phi(t)} = -\frac{t}{\Phi(t)} \int_{-\infty}^{t} \frac{\mathrm{e}^{-s^2/2}}{\sqrt{2\pi} s^2} \mathrm{d}s
$$

and $t + \Phi'(t)/\Phi(t) \geq 0$ for all $t \in \mathbb{R}$. Let $t < 0$ and $s = -t > 0$. We have $\Phi(t) = \bar{\Phi}(s) = \operatorname{erfc}(s/\sqrt{2})/2$ where $\operatorname{erfc} : \mathbb{R} \to \mathbb{R}_+$ is the complementary error function defined for all $u \in \mathbb{R}$ by $\operatorname{erfc}(u) = (2/\sqrt{\pi}) \int_u^{+\infty} \mathrm{e}^{-v^2} \mathrm{d}v$. By [GR14, Section 8.25, formula 8.254], we have the following asymptotic expansion for $s \to +\infty$

$$
\bar{\Phi}(s) = \frac{\mathrm{e}^{-s^2/2}}{\sqrt{2\pi} s} \left(1 - s^{-2} + 3s^{-4} + O(s^{-6}) \right) .
$$

Using that $\Phi'(t) = (2\pi)^{-1/2} \mathrm{e}^{-t^2/2}$ for all $t \in \mathbb{R}$, we get asymptotically for $t \to -\infty$ and $s = -t \to +\infty$,

$$
\Phi'(t)/\Phi(t) = s \left(1 + s^{-2} - 2s^{-4} + O(s^{-6}) \right) \tag{84}
$$

and $\lim_{t \to -\infty} h''(t) = -1$. There exists then $C > 0$ such that for all $t \in \mathbb{R}$, $-C \leq h''(t) \leq 0$. U_{pro} is then strongly convex, gradient Lipschitz and satisfies **H**2, **H**4, **H**1(k) for all $k \in \mathbb{N}$ and **H**5. By (84), we have for $t \to -\infty$ and $s = -t \to +\infty$, $h^{(3)}(t) = O(s^{-1})$. U_{pro} satisfies then **H**6.

C.4 Proof of Lemma 11

The proof is adapted from [FHS15, Lemma 1]. Let $i \in \{0, \ldots, 6\}$, $\varphi \in \mathrm{C}^{4+i}_{\text{poly}}(\mathbb{R}^d, \mathbb{R})$, $\bar{\gamma} > 0$, $\gamma \in [0, \bar{\gamma}]$ and $x, y \in \mathbb{R}^d$. Note that $\tau^{\text{MALA}}_{\gamma}(x, y)$ defined in (43) may be expressed as

$$\tau_\gamma^{\mathrm{MALA}}(x,y) = U(y) - U(x) - (1/2)\langle y-x, \nabla U(x) + \nabla U(y)\rangle \\ + (\gamma/4)\left\{\|\nabla U(y)\|^2 - \|\nabla U(x)\|^2\right\} . \quad (85)$$

A Taylor expansion of U and ∇U around x yields

$$U(y) - U(x) = \langle \nabla U(x), y-x\rangle + (1/2)\,\mathrm{D}^2 U(x)[(y-x)^{\otimes 2}] + (1/6)\,\mathrm{D}^3 U(x)[(y-x)^{\otimes 3}] \\ + (1/6)\int_0^1 (1-t)^3\,\mathrm{D}^4 U((1-t)x+ty)[(y-x)^{\otimes 4}]\mathrm{d}t , \quad (86)$$

$$\nabla U(y) = \nabla U(x) + \mathrm{D}^2 U(x)[y-x] + (1/2)\,\mathrm{D}^3 U(x)[(y-x)^{\otimes 2}] \\ + (1/2)\int_0^1 (1-t)^2\,\mathrm{D}^4 U((1-t)x+ty)[(y-x)^{\otimes 3}]\mathrm{d}t . \quad (87)$$

Substituting (86) and (87) into (85), we obtain for $z \in \mathbb{R}^d$, $\tau_\gamma^{\mathrm{MALA}}(x, x-\gamma\nabla U(x) + \sqrt{2\gamma}z) = \gamma^{3/2}\xi_\gamma(x,z)$ where ξ_γ is defined for all $x, z \in \mathbb{R}^d$ and $\gamma \in [0,\bar\gamma]$ by

$$\xi_\gamma(x,z) = -(1/12)\,\mathrm{D}^3 U(x)[(-\sqrt{\gamma}\nabla U(x) + \sqrt{2}z)^{\otimes 3}] \\ - (\sqrt{\gamma}/12)\int_0^1 (1-t)^2(1+2t)\,\mathrm{D}^4 U(x - t\gamma\nabla U(x) + t\sqrt{2\gamma}z)[(-\sqrt{\gamma}\nabla U(x)+\sqrt{2}z)^{\otimes 4}]\mathrm{d}t \\ + (1/2)\left\langle \nabla U(x), \int_0^1 \mathrm{D}^2 U(x - t\gamma\nabla U(x) + t\sqrt{2\gamma}z)[-\sqrt{\gamma}\nabla U(x) + \sqrt{2}z]\mathrm{d}t \right\rangle \\ + (\sqrt{\gamma}/4)\left\| \int_0^1 \mathrm{D}^2 U(x - t\gamma\nabla U(x) + t\sqrt{2\gamma}z)[-\sqrt{\gamma}\nabla U(x) + \sqrt{2}z]\mathrm{d}t \right\|^2 .$$

Note that by the dominated convergence theorem, for all $x, z \in \mathbb{R}^d$, $\lim_{\gamma\to 0}\xi_\gamma(x,z) = \xi_0(x,z)$ where ξ_0 is defined in (45). By (42), we get

$$R_\gamma^{\mathrm{MALA}}\varphi(x) - \varphi(x) = \mathbb{E}\left[\varphi(x - \gamma\nabla U(x) + \sqrt{2\gamma}Z) - \varphi(x)\right] \\ + \mathbb{E}\left[\left(\mathrm{e}^{-\gamma^{3/2}\xi_\gamma(x,Z)_+} - 1\right)\left\{\varphi(x - \gamma\nabla U(x) + \sqrt{2\gamma}Z) - \varphi(x)\right\}\right] , \quad (88)$$

where Z is an i.i.d. standard d-dimensional Gaussian variable. Combining (88) with the Taylor expansion (33), we get (i) with $\mathscr{A}_\gamma^{\mathrm{MALA}} : \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R}) \to \mathrm{C}^{i}_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$ given for all $\varphi \in \mathrm{C}^{4+i}_{\mathrm{poly}}(\mathbb{R}^d,\mathbb{R})$, $x \in \mathbb{R}^d$ and $\gamma \in (0,\bar\gamma]$ by

$$\mathscr{A}_\gamma^{\mathrm{MALA}}\varphi(x) = \mathscr{A}_\gamma^{\mathrm{ULA}}\varphi(x) + \mathbb{E}\Bigg[\gamma^{-3/2}\left\{1 - \mathrm{e}^{-\gamma^{3/2}\max(0,\xi_\gamma(x,Z))}\right\} \\ \times \left\{\int_0^1 \left\langle \nabla\varphi(x - t\gamma\nabla U(x) + t\sqrt{2\gamma}Z), \sqrt{\gamma}\nabla U(x) - \sqrt{2}Z\right\rangle \mathrm{d}t\right\}\Bigg] \quad (89)$$

and $\mathscr{A}_\gamma^{\mathrm{ULA}}$ given in (34). The assertion (ii) follows from taking the limit $\gamma \downarrow 0^+$ in (89) and the dominated convergence theorem.

Table 2: Estimates of the asymptotic variances for ULA, MALA and RWM and each parameter x_i, x_i^2 for $i \in \{1, \dots, d\}$, and of the variance reduction factor (VRF) on the example of the probit regression.

		MCMC	CV-1-MCMC		CV-2-MCMC		ZV-1-MCMC		ZV-2-MCMC	
		Variance	VRF	Variance	VRF	Variance	VRF	Variance	VRF	Variance
x_1	ULA	2.1	24	0.089	2.9e+03	0.00073	20	0.11	2.7e+03	0.00078
	MALA	0.41	22	0.019	2.7e+03	0.00015	18	0.023	2.6e+03	0.00016
	RWM	1.2	23	0.05	2.2e+03	0.00054	21	0.056	2.2e+03	0.00053
x_2	ULA	27	24	1.1	2.8e+03	0.0099	18	1.5	2.4e+03	0.011
	MALA	6.4	24	0.27	2.9e+03	0.0022	19	0.34	2.6e+03	0.0025
	RWM	13	18	0.72	1.8e+03	0.0073	16	0.81	1.8e+03	0.0075
x_3	ULA	11	24	0.47	6.7e+03	0.0017	18	0.62	6.3e+03	0.0018
	MALA	2.6	23	0.11	7e+03	0.00037	18	0.14	6.8e+03	0.00038
	RWM	5.5	18	0.3	4.3e+03	0.0013	16	0.34	4.3e+03	0.0013
x_1^2	ULA	0.75	3.5	0.22	1.6e+02	0.0048	2.8	0.26	1.3e+02	0.0057
	MALA	0.15	3.5	0.043	1.5e+02	0.001	2.8	0.053	1.3e+02	0.0011
	RWM	0.43	2.6	0.16	1.2e+02	0.0035	2.4	0.18	1.2e+02	0.0037
x_2^2	ULA	4.7e+02	9.3	51	1.4e+03	0.33	7.5	63	1.2e+03	0.4
	MALA	1.1e+02	9.1	12	1.5e+03	0.073	7.6	14	1.3e+03	0.085
	RWM	2.2e+02	7.7	29	1e+03	0.22	6.9	33	9.8e+02	0.23
x_3^2	ULA	1.1e+02	9.8	11	9.7e+02	0.11	7.9	14	7.9e+02	0.14
	MALA	24	9.7	2.5	9.8e+02	0.025	8.1	3	8.5e+02	0.029
	RWM	52	7.9	6.7	6.1e+02	0.086	7.1	7.4	5.9e+02	0.088

D Numerical experiments - additional results

We provide additional plots for the logistic regression, see Figure 2 and Figure 3, and the results for the Bayesian probit regression presented in Section 4, see Table 2, Figure 4 and Figure 5. They are similar to the results obtained for the Bayesian logistic regression.

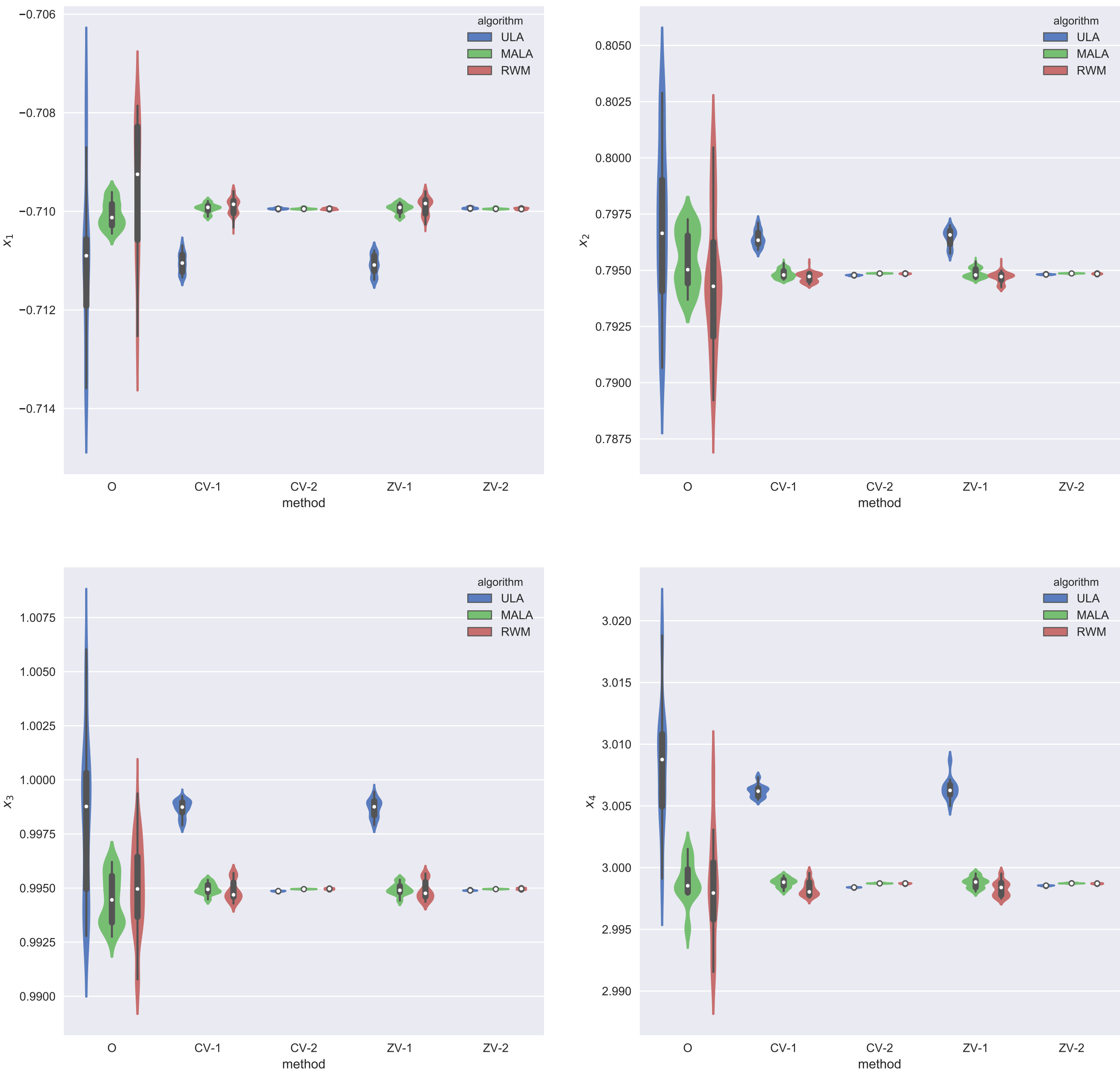

Figure 2: Boxplots of x_1, x_2, x_3, x_4 using the ULA, MALA and RWM algorithms for the logistic regression. The compared estimators are the ordinary empirical average (O), our estimator with a control variate (17) using first (CV-1) or second (CV-2) order polynomials for ψ, and the zero-variance estimator of [PMG14] using a first (ZV-1) or second (ZV-2) order polynomial basis.

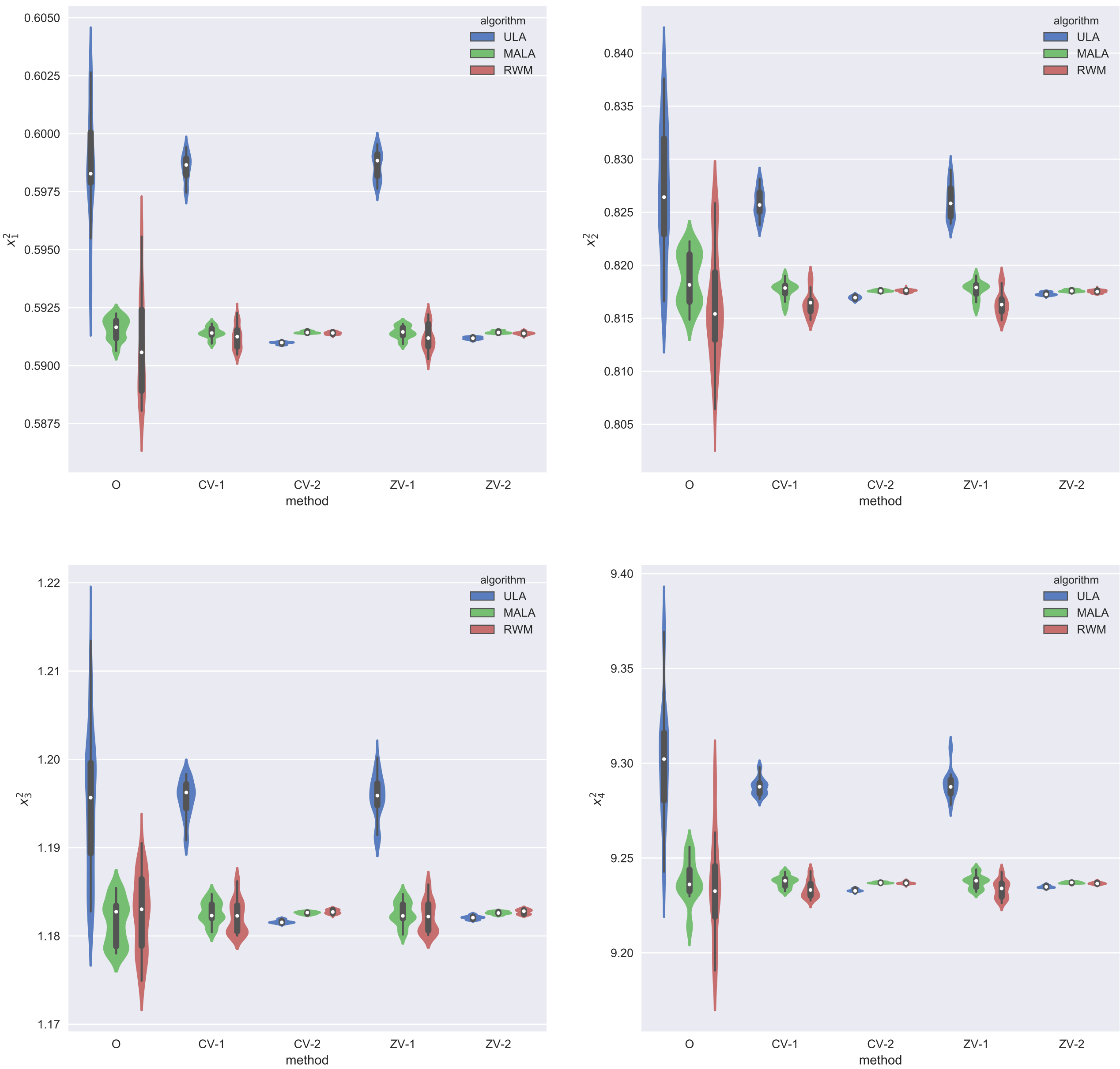

Figure 3: Boxplots of $x_1^2, x_2^2, x_3^2, x_4^2$ using the ULA, MALA and RWM algorithms for the logistic regression. The compared estimators are the ordinary empirical average (O), our estimator with a control variate (17) using first (CV-1) or second (CV-2) order polynomials for ψ, and the zero-variance estimator of [PMG14] using a first (ZV-1) or second (ZV-2) order polynomial basis.

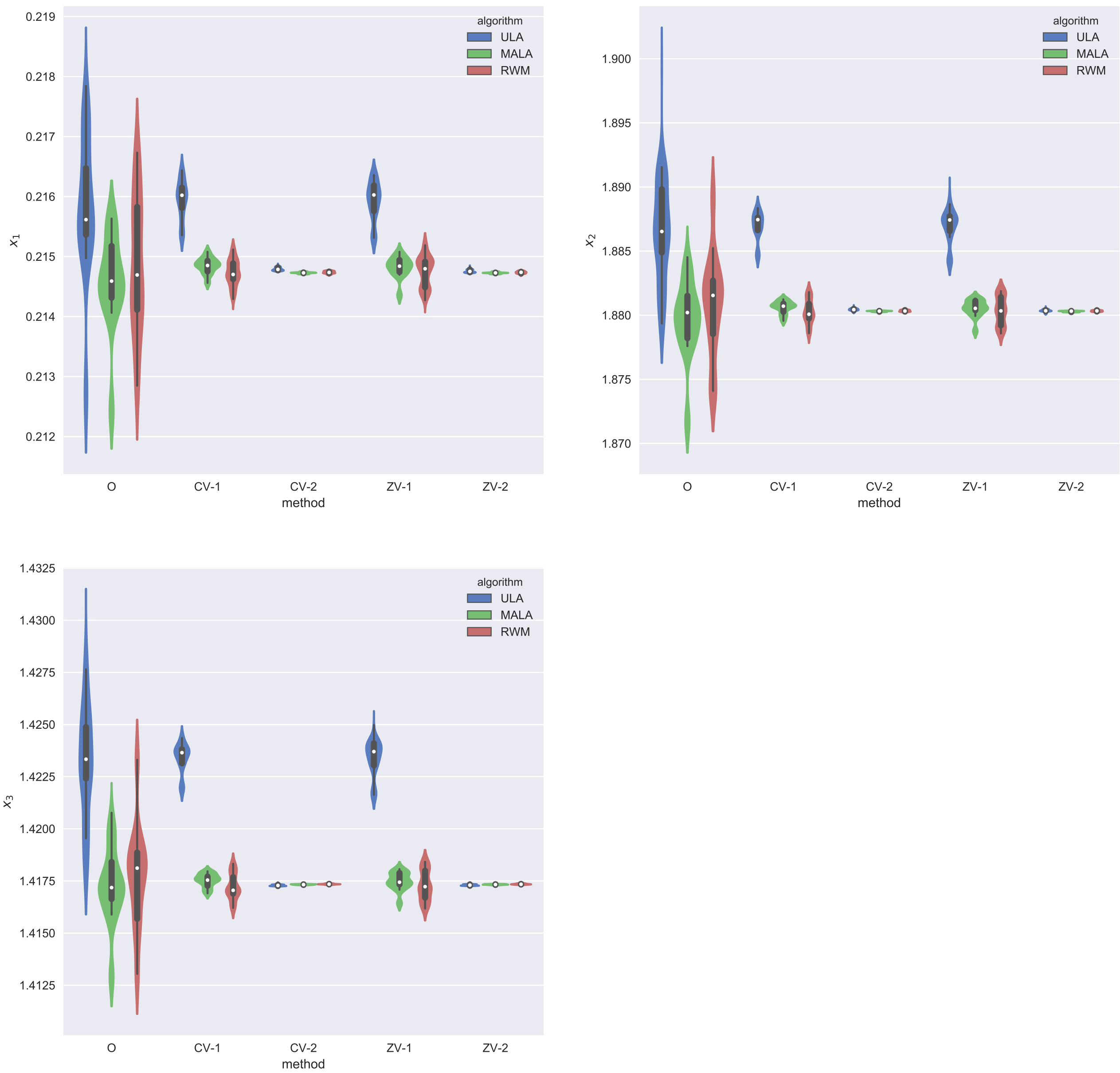

Figure 4: Boxplots of x_1, x_2, x_3 using the ULA, MALA and RWM algorithms for the probit regression. The compared estimators are the ordinary empirical average (O), our estimator with a control variate (17) using first (CV-1) or second (CV-2) order polynomials for ψ, and the zero-variance estimator of [PMG14] using a first (ZV-1) or second (ZV-2) order polynomial basis.

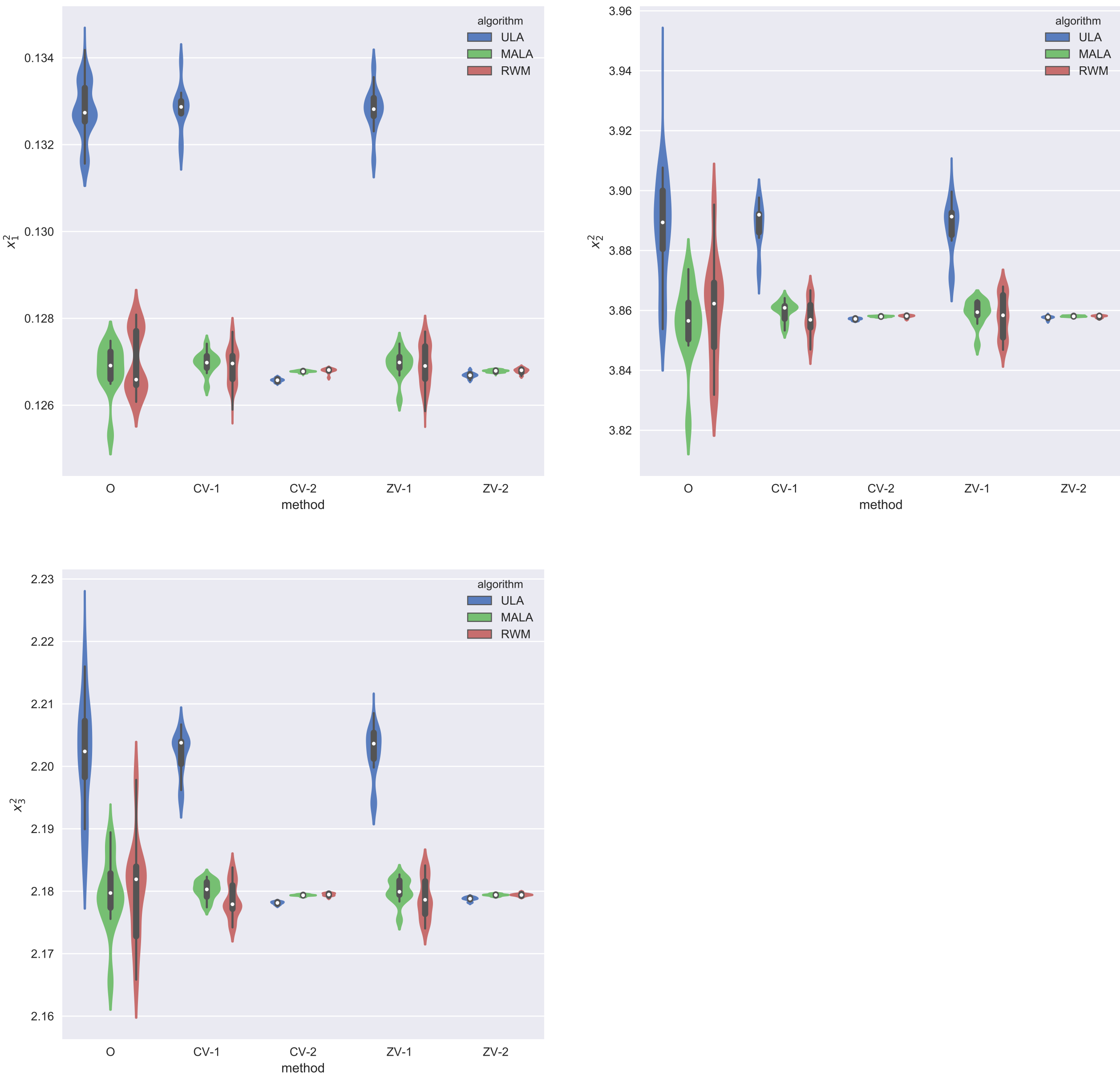

Figure 5: Boxplots of x_1^2, x_2^2, x_3^2 using the ULA, MALA and RWM algorithms for the probit regression. The compared estimators are the ordinary empirical average (O), our estimator with a control variate (17) using first (CV-1) or second (CV-2) order polynomials for ψ, and the zero-variance estimator of [PMG14] using a first (ZV-1) or second (ZV-2) order polynomial basis.